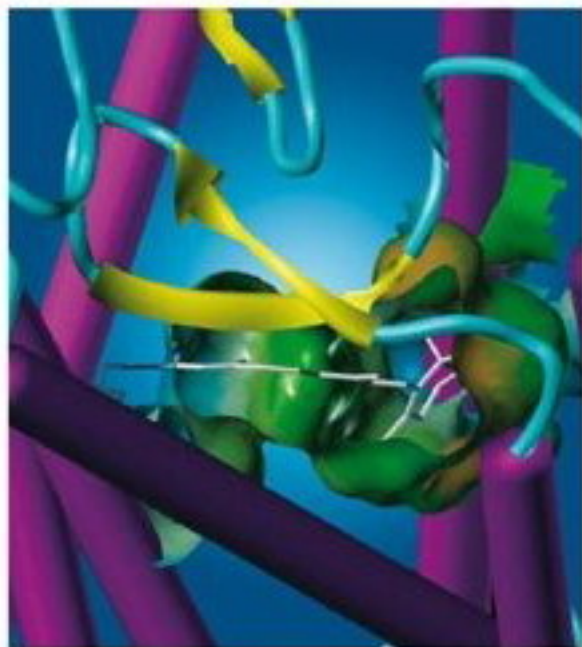Edited by Didier Rognan

⑭WILEY-VCH

Ligand Design for G Protein-coupled Receptors

Volume 30

Series Editors:
R. Mannhold,
H. Kubinyi,
G. Folkers

Ligand Design
for G Protein-coupled Receptors

Edited by
Didier Rognan

Methods and Principles in Medicinal Chemistry

Edited by R. Mannhold, H. Kubinyi, G. Folkers

Editorial Board
H.-D. Höltje, H. Timmerman, J. Vacca, H. van de Waterbeemd, T. Wieland

Previous Volumes of this Series:

Th. Dingermann, D. Steinhilber,
G. Folkers (eds.)

**Molecular Biology
in Medicinal Chemistry**

Vol. 21

2004, ISBN 3-527-30431-2

H. Kubinyi, G. Müller (ed.)

**Chemogenomics
in Drug Discovery**

Vol. 22

2004, ISBN 3-527-30987-X

T. I. Oprea (ed.)

**Chemoinformatics
in Drug Discovery**

Vol. 23

2005, ISBN 3-527-30753-2

R. Seifert, T. Wieland (eds.)

**G-Protein Coupled Receptors
as Drug Targets**

Vol. 24

2005, ISBN 3-527-30819-9

O. Kappe, A. Stadler

**Microwaves in Organic
and Medicinal Chemistry**

Vol. 25

2005, ISBN 3-527-31210-2

W. Bannwarth, B. Hinzen (eds.)

Combinatorial Chemistry

Vol. 26

2005, ISBN 3-527-30693-5

G. Cruciani (ed.)

Molecular Interaction Fields

Vol. 27

2005, ISBN 3-527-31087-8

M. Hamacher, K. Marcus, K. Stühler, A. van Hall,
B. Warscheid, H. E. Meyer (eds.)

Proteomics in Drug Design

Vol. 28

2005, ISBN 3-527-31226-9

D. J. Triggle, M. Gopalakrishnan, D. Rampe,
W. Zheng

**Voltage-Gated Ion Channels
as Drug Targets**

Vol. 29

2006, ISBN 3-527-31258-7

Ligand Design
for G Protein-coupled Receptors

Edited by
Didier Rognan

WILEY-VCH

WILEY-VCH Verlag GmbH & Co. KGaA

Series Editors

Prof. Dr. Raimund Mannhold
Biomedical Research Center
Molecular Drug Research Group
Heinrich-Heine-Universität
Universitätsstrasse 1
40225 Düsseldorf
Germany
Raimund.mannhold@uni-duesseldorf.de

Prof. Dr. Hugo Kubinyi
Donnersbergstrasse 9
67256 Weisenheim am Sand
Germany
kubinyi@t-online.de

Prof. Dr. Gerd Folkers
Collegium Helveticum
STW/ETH Zürich
8092 Zürich
Switzerland
folkers@collegium.ethz.ch

Volume Editor

Prof. Dr. Didier Rognan
Laboratoire de Pharmacochimie
de la Communication Cellulaire
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
67412 Illkirch
France

Contents

A Personal Foreword

Describing in a single book all existing approaches to design ligands targeting G protein-coupled receptors (GPCRs) is an impossible challenge. However, giving some clues to assist drug designers in their daily work is feasible. This is precisely the aim of the current book whose contributors have been selected to reflect the current knowledge on an extraordinary diverse family of protein targets. We have chosen to address fundamental and methodological issues which are the most likely to promote rational design of GPCR ligands. Ligand selectivity cannot be answered without considering the entire protein family at a genomic level. Dissecting the fine molecular mechanisms underlying GPCR function is also necessary to design ligands with the desired pharmacological profile. Last, a precise knowledge of current biostructural data is necessary to decide whether a ligand-based and/or a receptor-based design strategy is the most adequate. Many design strategies are indeed possible. Their potential is however very dependent on the current knowledge about a particular target and their ligands. It is therefore of outmost importance to be aware of all available information and design methods while beginning a drug discovery program. We do hope that this book will provide the reader with the necessary material to start with.

I would like to thank all contributors for their nice collaboration and their effort to respect a few deadlines in submitting their chapters. The series editors are also warmly acknowledged for their valuable comments in the early definition of the Table of Contents. I am grateful to Frank Weinreich and Renate Doetzer from Wiley-VCH for the nice collaboration over several months. Last, I extend my thanks to my family for their continuous support.

Illkirch, January 2006   *Didier Rognan*

Preface

G protein-coupled receptors (GPCR) represent to the best of our knowledge more at least 60 % of all receptors. This vast majority keeps them still alive as the most interesting group of targets in drug finding and development. Some 18.000 reviews are listed in Pubmed, many of them dealing with structural features and peculiarities of G protein-coupled receptors. Especially their functional categorization, association with other membrane-integral proteins and dimerization/oligomerization behaviour is still a hot topic in research.

Nevertheless, the existing body of knowledge at atomic resolution, enables us to propose interaction mechanism and activation models for this type of receptor. Here it is the merit of Didier Rognan, himself being on of the leading figures in the field of molecular modelling of GPCRs, that he started to collect a number of reputed researches sharing a history in the topic of GPCRs and edited a 12 chapter volume on the state-of-the-art in ligand design for those targets.

The volume starts with a genomic overview on GPCRs, which is followed by an appropriate review of the available data and their appearance and utilisation in databases. In more specialized chapters the question is raised how to de-orphanize receptors. Strategies in these fields are urgently needed since by HTS strategies, array technologies, etc., the number of orphan receptors has grown exponentially.

Ligand interaction does not mean at all that a drug will emerge from this knowledge. So, druggability analysis, which has overcome its infant years of rule-based estimates, has become a sophisticated methodology on its own. One chapter is devoted to druggability of human GPCRs. It's the molecular mechanism which is illuminated in depth within the subsequent three chapters. Oligomerization or just dimerzation, activation/inactivation processes and allosteric regulation are still complex puzzles to solve, last but not least because of the difficulties of understanding the entropy contribution.

Further chapters are dedicated to computational procedures. Chemical genomics approaches are going to be presented, the development detection of targeted libraries and priviledged structures for GPCR interaction and laedhopping and virtual screening approaches to ligand design.

The final three chapters deal with the 3D-structures of GPCRs and the usefulness as a basis for rational design of ligands. Both, modelling approaches as well as virtual screening will be discussed in extenso.

Thus, we expect this new volume in the series to be of fundamental interest to a large community of scientists and researches devoted to GPCRs. The editors are deeply convinced that the contents of this book will help to fathom the potential of GPCRs and will generate new ideas and visions for their role in drug discovery.

The editors are indepted to Renate Doetzer and Frank Weinreich from Wiley-VCH for their invaluable support in this project which is hereby gratefully acknowledged.

Raimund Mannhold, Düsseldorf
Hugo Kubinyi, Weisenheim am Sand
Gerd Folkers, Zürich

List of Contributors

Mohammad Alijaniaram
Centre for Addiction
and Mental Health
250 College St.
Toronto, Ontario M5T 1R8
Canada

Konstantin V. Balakin
ChemDiv, Inc.
11558 Sorrento Valley
Road Ste. 5
San Diego, CA 92121
USA

Frank Blaney
Department of Computational,
Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research
NFSP (North), Third Avenue
Harlow, Essex CM19 5AW
United Kingdom

Robert P. Bywater
Adelard Institute
9 Limes Avenue
London NW7 3NY
and
14 TM Ltd.
100 Low Road
Burwell, Cambridge CB5 OEJ
United Kingdom

Mercedes Campillo
Laboratori de Medicina Computacional
Unitat de Bioestadística and Institut de
Neurociències
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain

Anna-Maria Capelli
Department of Computational,
Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research
Via A. Fleming 4
37135 Verona
Italy

David Clark
Argenta Discovery Ltd.
8/9 Spire Green Centre
Flex Meadow
Harlow, Essex CM19 5TR
United Kingdom

Paul Denny-Gouldson
14 TM Ltd
100 Low Road
Burwell, Cambridge CB5 OEJ
United Kingdom

Xavier Deupi
Laboratori de Medicina Computacional
Unitat de Bioestadística and Institut
de Neurociències
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain

Robert Fredriksson
Department of Neuroscience
Uppsala University
Husargatan 3, Box 593
75124 Uppsala
Sweden

Jean-Luc Galzi
Faculté de Pharmacie
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
74, route du Rhin
67401 Illkirch
France

Susan R. George
Department of Pharmacology
and Medicine
University of Toronto
Medical Science Building
1 King's College Circle
Toronto, Ontario M5S 1A8
Canada

Cyril Goudet
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Cedric Govaerts
Institut de Recherche Interdisciplinaire
en Biologie Humaine et Moléculaire
Campus Erasme
Université Libre de Bruxelles
1070 Brussels
Belgium

Muriel Hachet-Haas
Département Récepteurs et Protéines
Membranaires
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
BP 10413
67412 Illkirch
France

Jacques Haiech
Département Récepteurs et Protéines
Membranaires
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
BP 10413
67412 Illkirch
France

Marcel Hibert
Faculté de Pharmacie
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
74, route du Rhin
67401 Illkirch
France

Christopher Higgs
ArgentaDiscovery Ltd.
8/9 Spire Green Centre
Flex Meadow
Harlow, Essex CM19 5TR
United Kingdom

XIX

Brigitte Ilien
Département Récepteurs et Protéines
Membranaires
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
BP 10413
67412 Illkirch
France

Xiaodong Ji
Centre for Addiction and Mental
Health
250 College St.
Toronto, Ontario M5T 1R8
Canada

Marie-Claude Kilhoffer
Faculté de Pharmacie
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
74, route du Rhin
67401 Illkirch
France

Thomas Klabunde
Sanofi-Aventis Deutschland GmbH
Science & Medical Affairs –
Drug Design
Industriepark Höchst G878
65926 Frankfurt
Germany

Sandra Lecat
Département Récepteurs et Protéines
Membranaires
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
BP 10413
67412 Illkirch
France

Emeline Maillet
Département Récepteurs et Protéines
Membranaires
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
Boulevard Sébastien Brant
BP 10413
67412 Illkirch
France

Tuan Nguyen
Centre for Addiction and Mental
Health
250 College St.
Toronto, Ontario M5T 1R8
Canada

Brian F. O'Dowd
Department of Pharmacology
and Medicine
University of Toronto
Medical Science Building
1 King's College Circle
Toronto, Ontario M5S 1A8
Canada

Leonardo Pardo
Laboratori de Medicina Computacional
Unitat de Bioestadística and
Institut de Neurosciènces
Universitat Autònoma de Barcelona
08193 Bellaterra
Spain

Jean-Philippe Pin
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Laurent Prézeau
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Didier Rognan
Laboratoire de Pharmacochimie
de la Communication Cellulaire
CNRS UMR 7175-LC1
Institut Gilbert Laustriat
74, route du Rhin
67401 Illkirch
France

Philippe Rondard
Département de Pharmacologie
CNRS UMR 5203 – INSERM U661
Université de Montpellier
141 rue de la Cardonille
34094 Montpellier cedex 5
France

Nikolay Savchuk
ChemDiv, Inc.
11558 Sorrento Valley
Road Ste. 5
San Diego, CA 92121
USA

Helgi B. Schiöth
Department of Neuroscience
Uppsala University
Husargatan 3, Box 593
75124 Uppsala
Sweden

Giovanna Tedesco
Department of Computational,
Analytical, and Structural Sciences
GlaxoSmithKline Medicine Research
Via A. Fleming 4
37135 Verona
Italy

Sergey E. Tkachenko
ChemDiv, Inc.
11558 Sorrento Valley
Road Ste. 5
San Diego, CA 92121
USA

1
G Protein-coupled Receptors in the Human Genome

Robert Fredriksson and Helgi B. Schiöth

1.1
Introduction

The superfamily of G protein-coupled receptors (GPCRs) is one of the largest families of proteins in the human genome [1, 2] and probably also in most other vertebrate species [3]. GPCRs participate in a diversity of important physiological functions and are targets for many modern drugs. Their ligands are particularly diverse, and include ions, organic odorants, amines, peptides, proteins, lipids, nucleotides and photons, which are all able to activate GPCRs. The main structural characteristic of GPCRs is seven stretches of about 25–35 consecutive amino acid residues that show a relatively high degree of hydrophobicity and represent α-helixes that span the plasma membrane in an anti-clockwise manner. These sequences stretch from the common area or a recognition and connection unit of all GPCRs, enabling an extracellular ligand to exert a specific effect on the cell. This area of the receptors is generally relatively well conserved and is used to identify and classify novel GPCRs as other areas of the receptors are frequently much more diverse. The name GPCRs indicates that these receptors interact with G-proteins. This has however not yet been demonstrated for most of the proteins classified as GPCRs. Moreover, GPCRs are known to have many alternative signaling pathways, interacting directly with a number of other proteins such as arrestins and kinases. Hence, it would perhaps be more technically correct to term this superfamily "seven transmembrane (TM) receptors", but the GPCR terminology has become more established.

Both physiological and structural features have been used to classify GPCRs. The first classification system was introduced in 1994 by Attwood and Findley [4]. They used the term "clans" to designate the different GPCR families. The classified dataset at this time contained over 240 rhodopsin-like GPCRs from different species. Many of these receptors were olfactory and light-recognizing receptors of the opsin type. Independently, but around the same time, Kolakowski presented the well known "Family A–F classification system" [5]. This system included receptors shown to bind G-proteins while the other 7TM receptors were classified as O (other). In conjunction with this classification system the database GPCRdb was

developed and included at that time 777 unique GPCRs from various species. Family A contained receptors similar to rhodopsin and biogenic amine receptors. Family B contained receptors similar to the secretin and calcitonin receptors while Family C contained the metabotropic glutamate receptors. Family D and E contained only receptors that were not, and still have not been, identified in mammals, namely the fungal pheromone receptors and the cAMP binding receptors, respectively. Finally, Family F contained archebacterial opsins. The Kolakowski classification system was later extended independently, and differently, by Josefsson and Flower in 1999 [6, 7]. Moreover, another classification system was suggested in 1999 that contained in total five families based on the position of the ligand-binding pocket and the sequence length of the receptors. This system excluded the receptors that are not present in vertebrates [8]. This system used both structural and physiological features to classify the receptors. Recently, we have undertaken large-scale systematic phylogenetic analyses including the majority of the GPCRs in the human genome [9]. This provides us with the GRAFS system showing five main families named *Glutamate* (G; previous family C/3), *Rhodopsin* (R; previous family A/1), *Adhesion* (A; previously part of family B/2), *Frizzled/Taste2* (F; previously O/5 and not included) and *Secretin* (S; previously part of family B/2). Moreover, we subdivided the large Rhodopsin family into 13 subgroups. The grouping was carried out using strict phylogenetic criteria and only a few human receptors did not group into these clusters and these receptors were thus placed into what we called *Other 7TM receptors*. There are several GPCRs that have been discovered since we published this classification [10–13] and here we present an updated version of the human repertoire. In this overview we describe each of the families and groups within the GRAFS classification system and include phylogenetic trees which were derived by Maximum Likelihood and show branch lengths.

1.2
The Adhesion Family

The Adhesion family is the second largest GPCR-family in humans with 33 members. The group is called Adhesion GPCRs according to a recent GPCR classification [9] and this nomenclature seems to prevail. This family has however been assigned various names through the years. These include EGF-TM7 to reflect the presence of epidermal growth factor (EGF) domains in the N-termini [14, 15] and LN-TM7 receptors where LN stands for long N-termini and B2/LNB-7TM to reflect their vague similarity to secretin receptors [16]. The Adhesion family members have several structural features that clearly separate them from all other groups of GPCRs. In a recent article we showed the entire repertoire in human and mouse where the diversity of their N-termini is highlighted [12]. Their long N-termini contain a high percentage of Ser and Thr residues that can create O- and N-glycosylation sites. These N-termini or stalk-like regions are thus thought to be highly glycosylated and act as a mucin-like domain with a rigid erect structure protrud-

ing from the cell surface. The long N-termini are believed to bind various proteins that promote cell-to-cell and cell-to-matrix interactions. All of the Adhesion GPCRs except GPR 123 contain a GPCR proteolytic domain (GPS). Additionally their N-termini can contain a number of different domains that are also found in various other proteins, such as cadherin, lectin, laminin, olfactomedin, immuno-globulin or trombospondin. It is likely that the repertoire of these domains plays an important role in the functional specificity of the receptors.

Phylogenetically, as can be seen in Fig. 1.1, this family forms three main subfamilies with the largest containing lectomedin, EGF-like module, cadherin EGF, EGF lathrophilin, CD97 and GPR 127 receptors. It is interesting to note that all receptors in this group, with the exception of lectomedin receptors, contain EGF domains and that no receptors outside this cluster contain this type of domain [12]. Continuing clockwise in Fig. 1.1 the second group contains 10 receptors termed GPR 110, GPR 111, GPR 113, GPR 115, GPR 116, GPR 123, GPR 124, GPR 125, GPR 133 and GPR 144. The receptors in this cluster have in general very few recognizable domains in their long N-termini, with GPR 123 having no known domains and GPR 110, GPR 111 and GPR 115 having only a GPS domain. The other receptors contain immunoglobulin domains (GPR 124, GPR 125 and GPR 116), hormone-binding domains (GPR 113), leucine-rich repeats (GPR 124 and GPR 125), a pentraxin domain (GPR 144) and a sea urchin sperm domain (GPR 116) [12]. The third family contains brain angiogenesis receptors, human epidymal receptors, the very large GPR 1 (over 6300 amino acids long) and GPR 56, GPR 97, GPR 112, GPR 114, GPR 126 and GPR 128. Also, several receptors in this group are rather sparse in known domains with GPR 56, GPR 97, GPR 114, GPR 126, GPR 128 and human epidymal receptor containing only GPS domains. The three Brain Angiogenesis Inhibitor GPCRs (BAI) contain only hormone-binding and trombospdin domains while GPR 112 contains a pentraxin domain. The very large GPR 1 contains several copies of the sodium–calcium exchange/integrin beta domains [12]. Although several of the more recently discovered Adhesion GPCRs have surprisingly few recognizable functional domains in their N-termini, it is likely that these receptors contain novel domains that are not recognizable using current bioinformatics tools. The majority of the Adhesion GPCRs are orphans and for the few that have been characterized with regard to ligand binding, none has been shown to bind their ligand within the TM regions. CD97 is one of the most studied receptors in this family and is found in several types of blood cell. CD97 interacts with the 312-amino acid membrane protein CD55 (or decay accelerating factor; DAF) which is expressed on most leukocytes [17]. Recently, it was also shown that a glycosaminoglycan (chondroitin sulfate) acts as a cellular ligand specific to the EGF-like domains of the EMR2 [15]. Receptors of the Adhesion family are expressed in various parts of the human body and many of them have prominent expression in the immune system, central nervous system, and in the reproductive organs, suggesting that they might take part in a large variety of physiological functions.
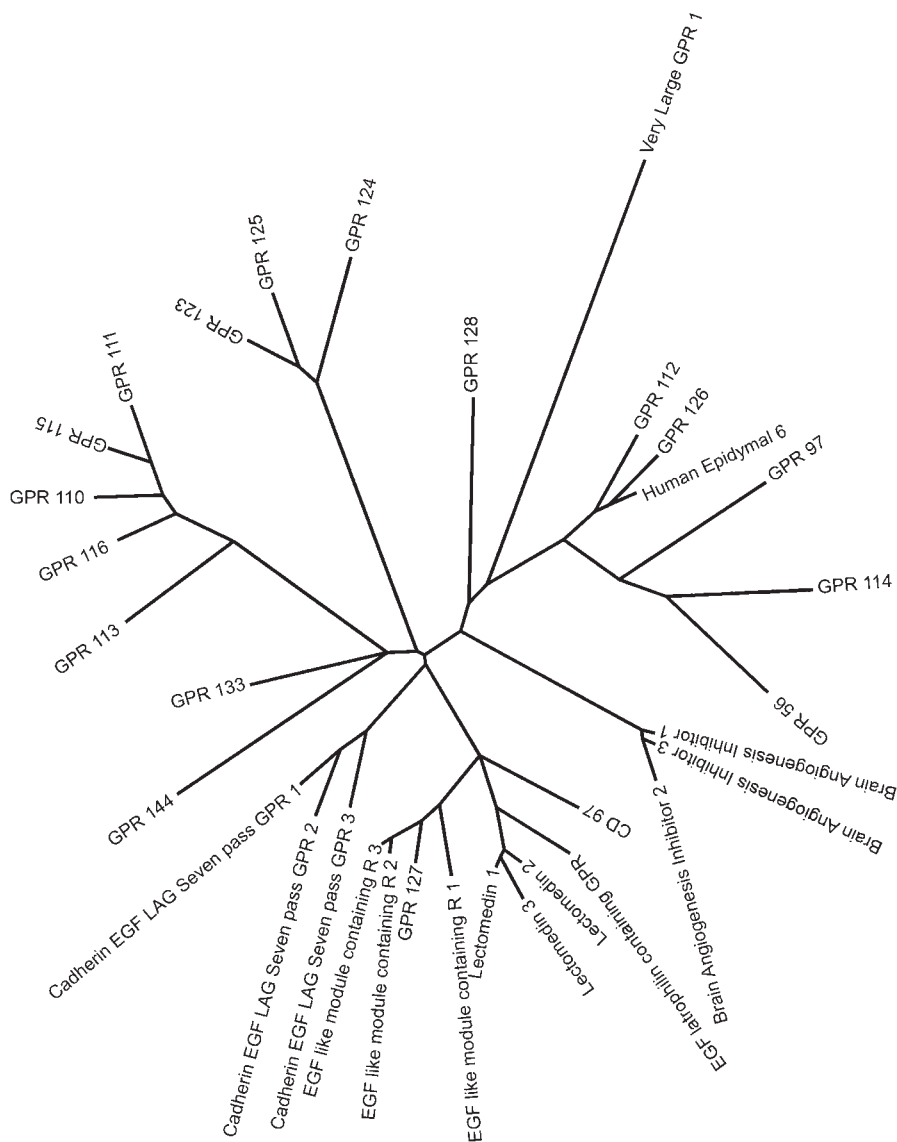
Fig. 1.1 Phylogenetic trees for each family and group of GPCRs. The topology is a consensus tree from 100 bootstrap replicas calculated using ordinary parsimony and the branch lengths are optimized using the Maximum Likelihood method with no assumptions of the presence of a molecular clock and hence the branch lengths correspond to protein distances, not evolutionary time. The calculations are performed on the amino acid sequences from transmembrane region 1 to the end of transmembrane region 7, meaning that no N-termini are used for the calculations as they are very divergent for the different groups. R is the abbreviation for receptors in the annotation on the figures.

1.3
The Secretin Family

The secretin family consists of 15 receptors and occurs widely in all animal species [16, 18]. The N-terminal regions of these receptors share some primary sequence similarity with the Adhesion family of GPCRs and in the A-F classification system both are considered to belong to the B family. Although this sequence similarity is clearly recognizable, it is evident that the Secretin and the Adhesion families are evolutionarily old and that they split into individual groups long ago. Both families are present as multimember families in both insects such as *D. melanogaster* and *A. gambiae* as well as in *C. elegans* [3] as are the other main families, i.e. the Adhesion, Rhodopsin, Frizzled and Glutamate families. The phylogenetic tree of this family has four main subgroups, the largest one consisting of the secretin, Growth Hormone Releasing Hormone, Vasoactive Intestinal Peptide and Pituitary Adenylate Cyclase-Activating Polypeptide (PACAP) receptors. The other groups contain, in clockwise order in Fig. 1.2, the Corticotropin Releasing Hormone/Calcitonin Gene Related Peptide Receptors, Glucagon/Glucagon-like Peptide/Gastric inhibitory peptide receptors and the Parathyroid Hormone Receptors. The receptors in the Secretin family bind rather large peptides and most often act in a paracrine manner. The Secretin family name is related to the fact that the secretin receptor was the first of this family to be cloned and the term secretin-like receptors has also frequently been used in the literature with reference to receptors in this cluster. The N-terminal, between about 60 and 80 amino acids long, contains conserved Cys bridges and is particularly important for the binding of the ligand to these receptors. For example, the N-terminal alone of the VIPR and PACAP receptor constitutes a functional binding site for the ligand. The receptors have a recognizable "hormone binding domain" in the N-termini and these receptors bind rather large peptides that most often act in a paracrine manner.

1.4
The Frizzled/Taste 2 Family

Our phylogenetic studies on the human repertoire have indicated that two very different groups of receptors cluster together. There are few elements in the consensus sequence and the HMM models, such as the consensus sequence of IFL in TM2, SFLL in TM5, and SxKTL in TM7 which are motifs that do not seem to be present in the consensus sequences of the other four families, that could explain why these two groups of receptors cluster together. Further studies are needed to investigate whether these two groups have a common evolutionary history. Below we look at these groups separately.
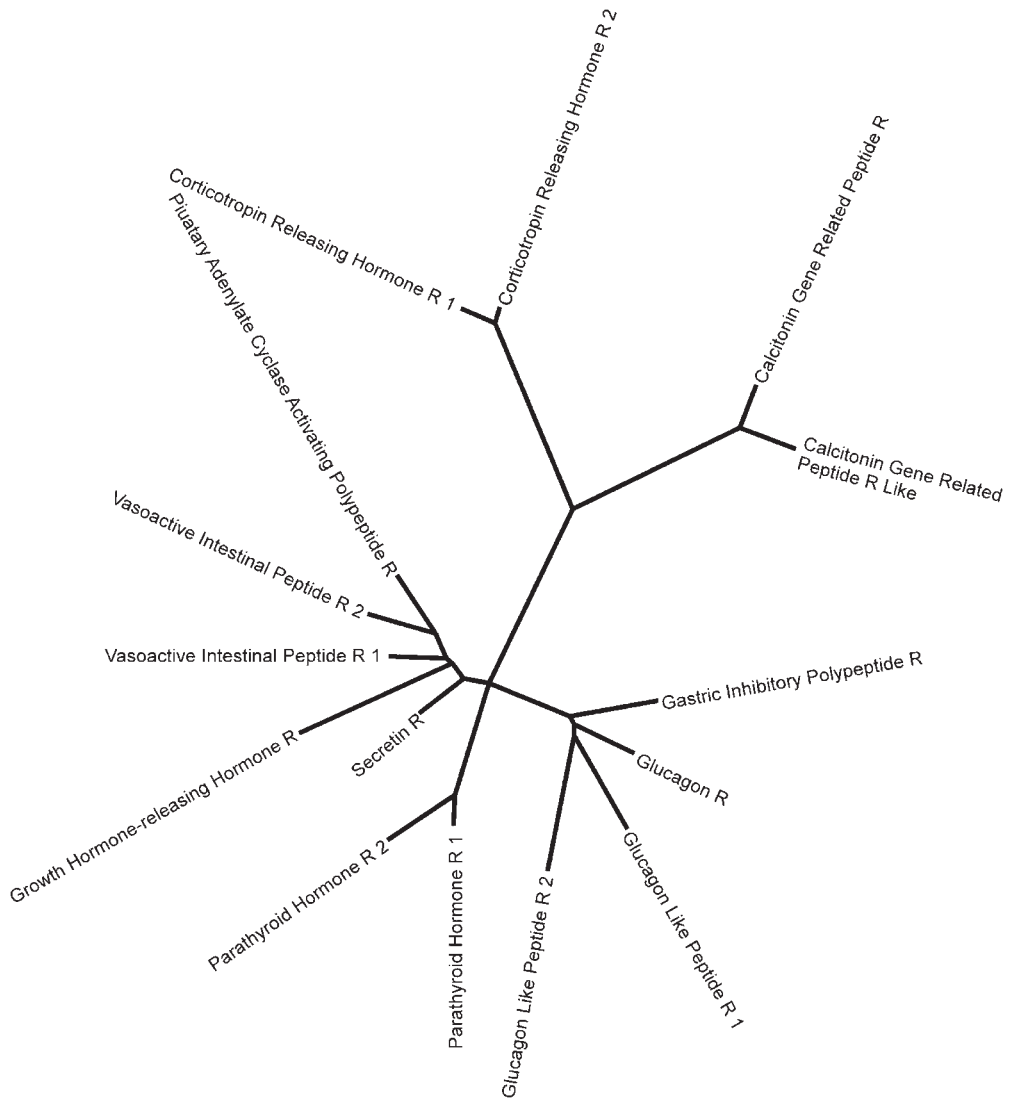
Fig. 1.2 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

1.4.1
The Frizzled Receptor Cluster

The Frizzled group consists of 10 frizzled receptors named Frizzled receptor 1–10 and the single Smoothened receptor. The topology of the tree in Fig. 1.3 shows four main clusters: the cluster containing Frizzled 1, 2 and 7 which have approximately 75% identity to each other; the Frizzled 8 and 5 that have 70% identity, the
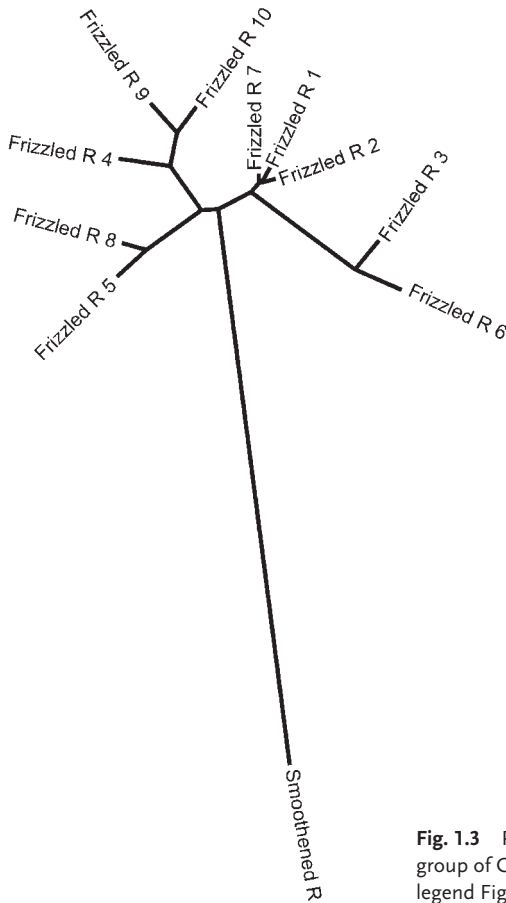
Fig. 1.3 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

Frizzled 10, 9 and 4 that have around 65% identity; and finally Frizzled 6 and 3 that have 50% amino acid identity. The identities shared by receptors from different clusters are between 20 and 40%, indicating that four parental genes from the Frizzled family were initially formed and subsequently the four clusters originated out of these. Smoothened is, as evident in Fig. 1.3, clearly the most divergent of the receptors from the Frizzled family, sharing only 24% amino acid identity to FZD2 and less to the others. The large evolutionary distance between the Smoothened receptor and the other Frizzled receptors also reflects a large evolutionary time of divergence as Smoothened are found as a distinct receptor back in *C. elegans*, which diverged from the lineage leading to mammals more than 600 million years ago [19]. Despite this large sequence divergence between Smoothened and the other Frizzled receptors, all these receptors clearly belong to the same family, which has been shown from phylogenetic analysis of the entire GPCR family [9]. The frizzled receptors control cell fate, proliferation, and polarity during metazoan development by mediating signals from secreted glycoproteins termed Wnt. The

frizzled name was first used for a receptor cloned from *Drosophila*, referring to the curled and twisted Wnt ligand. It was for some time questionable whether the Frizzled receptors were actually true GPCRs, but it has been shown that the Wnt ligand binds to the rat Frizzled receptor 2 and can induce G-protein coupling [20] providing evidence that the frizzled proteins are GPCRs. The frizzled family of receptors has about 200 amino acid-long N-termini with conserved cysteines that are likely to participate in Wnt binding.

1.4.2
The Taste 2 Receptor Cluster

The Taste receptors type 2 (T2Rs) are GPCRs without introns and have very short extracellular N-termini which are believed to be unable to bind ligands and hence it has been suggested that the ligands bind in a pocket within the extracellular parts of the TM regions [21]. The T2Rs show very low sequence similarity to the umami and sweet taste receptors within the Glutamate family (see below), which indicates that the function of recognizing the taste of substances has undergone at least two developmental stages in animals. T2Rs recognizes bitter substances and the relatively large number of T2Rs suggests that mammals have the capability of recognizing many different bitter-tasting substances. Because many poisonous compounds have a bitter taste it has been suggested that this large repertoire of bitter taste receptors has evolved as a key defense mechanism [22]. To date, the majority of T2Rs are orphan receptors without known identified ligands and the only human receptor with a known ligand is T2R16, which has been shown to be activated by salicin [23]. It has also been shown that two mouse and one rat receptor can be activated by bitter compounds and it is highly plausible that the other T2Rs also respond to bitter compounds [22]. Phylogenetically (Fig. 1.4) the relationship between the T2Rs varies considerably with both long branches and clusters of receptors with relatively short branches, such as the cluster of nine receptors containing Taste 2R62. As Taste receptors type 2 are specific for vertebrates [3] and hence have only been around for maybe 450 million years, this suggests that this family is evolving rapidly, perhaps the most rapidly evolving among all GPCR groups. Additionally, it is likely that the clusters of receptors with short branches have arisen rather recently and this is also supported by the fact that these receptors are located in a gene cluster on human chromosome 12 (Bjarnadottir et al., unpublished data).

1.5
The Glutamate Family

The Glutamate family, also known as the clan C receptors, is mostly known for the metabotropic glutamate receptors, which for example mediate glutamate responses in a variety of CNS functions. Other subgroups within this family are the GABA-receptors, the calcium-sensing receptors, a number of orphan GPCRs and the chemosensory receptors, more specifically taste receptors type 1 (TAS1R) and
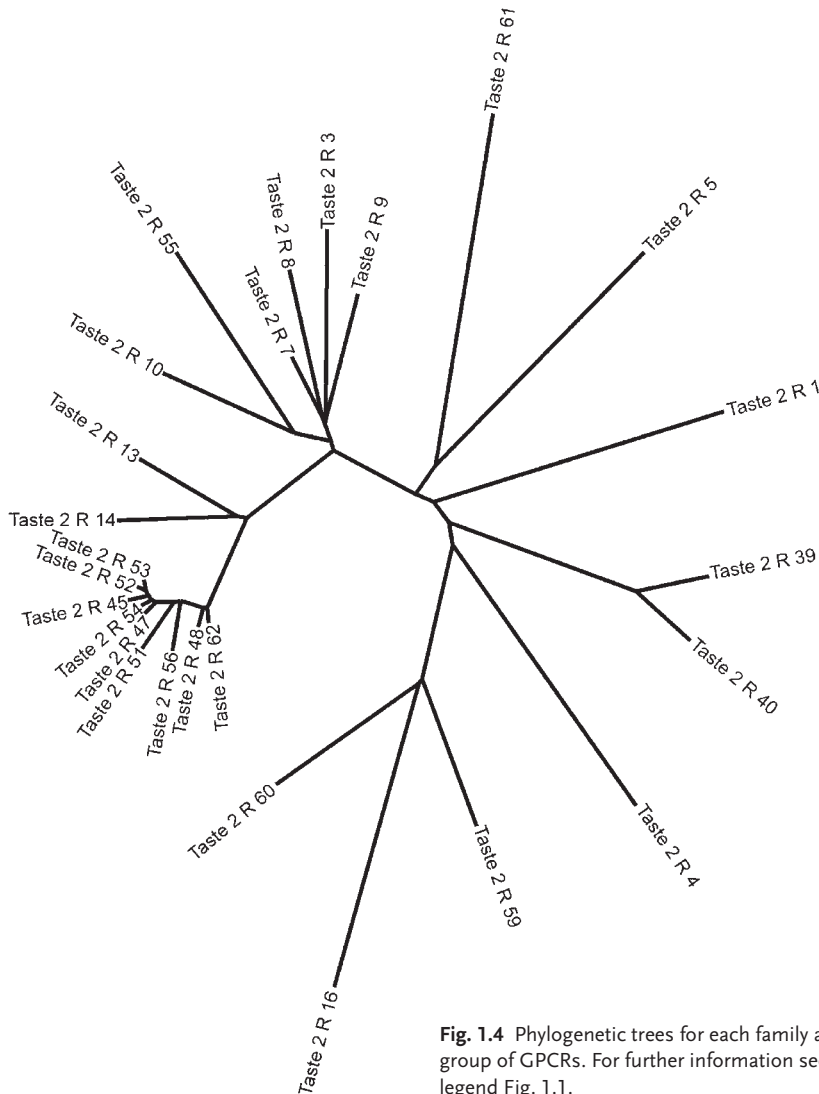
Fig. 1.4 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

the G_0-coupled pheromone receptors (V2Rs; see Fig. 1.5). The V2Rs are specialized in the detection of pheromones related to social and reproductive behavior in most terrestrial vertebrates [24]. Although V2Rs are present in large numbers in all non-primate mammalian genomes investigated, no functional V2Rs are present in the human genome. TAS1R have been identified in mouse, rat and human and are activated by sweet and amino acid taste compounds and these receptors show no close evolutionary relationship to the bitter taste receptors and it is probable that these two families of taste receptors have arisen independently. Surprisingly though, both these types of taste receptors appear to be present only in verte-
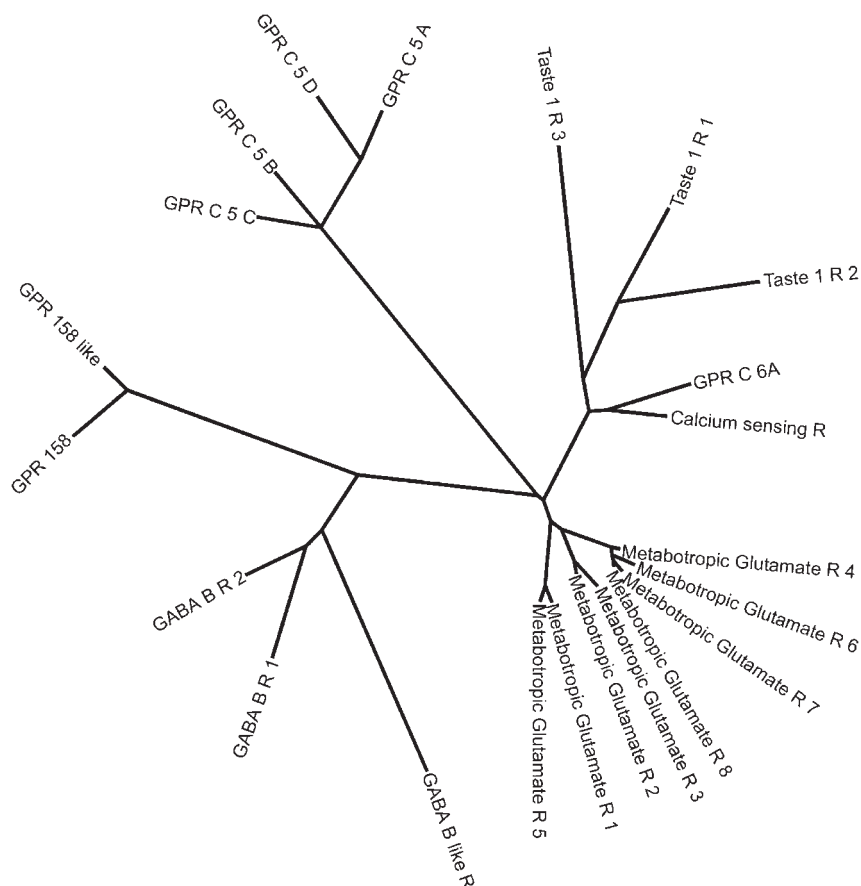
Fig. 1.5 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

brates [3] and might hence have evolved independently but at approximately the same time. This functional diversity of receptors belonging to the same phylogenetic group, the Glutamate GPCRs, is seemingly only matched by the large Rhodopsin family. One of the most important structural features of many of the Glutamate GPCRs is that they possess a large extracellular domain, usually around 600 amino acids. Two-thirds of the extracellular domain has been shown to participate in a two-lobe "Venus flytrap" mechanism which is critical for recognizing and binding ligands among the subgroups of metabotropic glutamate-, GABA- and calcium-sensing receptors. Also the taste receptors and the orphan receptors related to GABA and calcium-sensing receptors have rather long extracellular N-termini. The Glutamate family contains 22 receptors in the human genome (Fig. 1.5) and these are phylogenetically divided into four main groups. The largest is the group of closely related metabotropic glutamate receptors which comprises eight members. Continuing clockwise along the tree, there is a group of

five receptors containing two GABA receptors and three orphans. It is evident that these receptors are clearly distantly related considering the branch lengths of the tree. The next group contains four orphan receptors, GPR C 5A–Ds, which are the only receptors in the Glutamate family that lack long N-termini. The fourth group contains the three taste receptors of type 1, the calcium-sensing receptor and the orphan receptor GPR C 6A.

1.6
The Rhodopsin Family

The Rhodopsin family is made up of the largest number of receptors with about 278 non-olfactory receptors. In addition there are at least 347 olfactory receptors [25] and another 11 receptors currently placed in the "Other" group that are likely to be of Rhodopsin type and this adds up to 636 known Rhodopsin GPCRs. The Rhodopsin family corresponds to what has previously been called either the rhodopsin-like receptors or clan A in the A–E classification system. The crystal structure of bovine rhodopsin has been revealed [26] and this is the only animal GPCR that has had its exact structure determined. Therefore bovine rhodopsin has frequently been used as a template for modeling the structure of other GPCRs from the rhodopsin family [27–29]. It should be noted that bacteriorhodopsin, which has also had its three-dimensional structure determined, has no sequence similarity with the GPCRs in the human genome [6]. The ligands for most of the rhodopsin receptors bind within a cavity between the TM regions [30]. There are however important exceptions to this, in particular for the glycoprotein binding receptors (LH, FSH, TSH and LG), where the ligand-binding domain is in the N-terminal. Our analysis showed four main groups [9] which we have designated α, β, γ and δ. Since our original publication we have identified another 15 non-olfactory rhodopsin GPCR [13] (Gloriam et al., unpublished data) and although these are clearly atypical most seem to belong to one of the four main groups.

1.6.1
The Rhodopsin α-Group

This group has five main branches namely the prostaglandin-, amine-, opsin-, melatonin-, and M.E.C.A. receptor cluster as can be seen in Fig. 1.6. This is the largest of the four main groups in the Rhodopsin family with 101 members in total.

1.6.1.1 The Prostaglandin Receptor Cluster
This cluster contains in total 15 receptors, the seven prostaglandin receptors and the closely related tromboxane receptor together with seven orphan receptors. The orphan receptors are divided into three groups: one containing the super conserved receptor expressed in the brain, one containing group GPR 61 and GPR 62 and one containing GPR 26 and GPR 78.

Fig. 1.6 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

1.6.1.2   The Amine Receptor Cluster

The Amine receptor cluster contains serotonin, dopamine, muscarinic, histamine, adrenergic, and trace amine receptors. Interestingly this large group contains only one orphan receptor, GPR 101. All the known ligands for the receptors in this group are structurally related small amine molecules with a single aromatic ring. The serotonin receptors display a heterogeneous phylogenetic pattern and two distinct subgroups can be seen, the serotonin receptor 1 and serotonin

receptor 2 branches. Serotonin receptor 5 and 7 cluster basally of the serotonin receptor 1 and 2 but the branches are relatively long which indicates that these receptors are still rather distantly related. Surprisingly serotonin receptors 4 and 6 do not cluster with the other serotonin receptors but rather basally in the Amine group. This indicates that the evolutionary history of the serotonin receptors is peculiar and complex and that the ability to bind the ligand serotonin may have arisen several times in an independent manner during evolution. The trace amine receptor branch contains the trace amine receptors 1 to 5, where trace amine receptor 2 is a pseudogene in humans and hence excluded, and the other four receptors are putative neurotransmitter receptors, also known to bind trace amines. These receptors are closely related to each other with regard to amino acid identity, which indicates that they have arisen recently. The repertoire of trace amine receptors is also very dynamic as species-independent gene family expansions have been shown to have taken place in rat (17 trace amine receptors), mouse (13 trace amine receptors) and zebrafish (57 trace amine receptors; Gloriam et al., unpublished data). The five muscarinic acetylcholine receptors form a homogenous cluster within the amine group, with a relatively high degree of sequence similarity. The adrenergic receptors A and B, also called adrenergic receptors α and β, form separate clusters branching basally in the Amine receptor cluster. This shows that the A and B adrenergic receptors are rather distantly related. The dopamine receptors on the other hand do not cluster together but rather fall into two groups containing two and three receptors each. One of these groups, containing dopamine receptors 2, 3 and 4 is located basally in the adrenergic A receptor branches and the group containing dopamine receptors 1 and 5 is positioned in the adrenergic B receptor branch. This prompts the speculation that dopamine and adrenergic receptors share a relatively recent evolutionary origin. Perhaps these two receptor families originate from two families consisting of promiscuous dopamine- and adrenalin-binding receptors and in each of these families specialization with regard to ligand preferences has evolved. One other heterogeneous group of receptors in the Amine cluster is the histamine receptors, three of these clusters have relatively long branches together with the orphan receptor GPR 101 while histamine receptor 2 is placed in a position basal to the trace amine receptors.

1.6.1.3 The Opsin Receptor Cluster

This cluster of Rhodopsin α-receptors comprises the rhodopsin receptor, the three visual cone pigments for long, short and medium wavelength photons, peropsin, encephalopsin, melanopsin, the retinal G-protein coupled receptor and GPR 136. The opsins for long and medium wavelength light are found in the same chromosomal position, Xq28 only 23 700 bases apart, as a result of a gene duplication specific for primates. Unequal crossing over between these genes, resulting in either loss of one of the genes or a hybrid (chimeric) version, is the cause of the commonest form of color blindness [31]. These two proteins are over 96 % identical at the amino acid level.

1.6.1.4 The Melatonin Receptor Cluster

The two melatonin receptors cluster on a branch of their own together with or-phan receptor GPR 50. Melatonin is a hormone that is mainly produced and se-creted at night by the pineal gland. GPR 50 has been linked with a sex-specific risk factor for susceptibility to bipolar disorder [32].

1.6.1.5 The MECA Receptor Cluster

This cluster consists of the melanocortin, endothelial differentiation sphingolipid, cannabinoid, and adenosine receptors. The cluster also contains the three recep-tors GPR 3, GPR 6 and GPR 12 that recently have been shown to bind lipid li-gands [33]. It is interesting to note that the receptors in this group, although clearly closely related phylogenetically, bind structurally very different ligands; melanocortin receptors bind peptides (melanocortins), endothelial differentiation sphingolipid receptors bind lysophosphatidic acid, cannabinoid receptors have anandamide (arachidonylethanolamide) as their endogenous ligand, GPR 3, 6 and 12 bind lipids and adenosine receptors bind adenosine (a purine sugar deriva-tive). The common feature of some of ligands in this group is that both lysopho-sphatidic acid anadamide and the lipids binding GPR 3, 6 and 12 are derivatives of phospholipids but the adenosine and peptide ligand-binding receptors in this cluster are still peculiar.

1.6.1.6 Other Rhodopsin α-Receptors

A number of more or less unrelated orphan receptors that clearly belong to the Rhodopsin α-cluster do not fall into any of the established groups of receptors. These are the related receptors GPR 45 and GPR 63, GPR 119, GPR 148, the two related GPR 162 and 163 receptors, GPR 84, GPR RE2 and the related receptors GPR 52 and 21. As evident from Fig. 1.6, although they figuratively appear to be related as they are positioned on a common branch, these receptors have long branches to the other receptors in the Rhodopsin α-group, even those on the same branch. Therefore the clustering, and hence the apparent relationship of these re-ceptors, might be a result of a phenomenon known as long-branch attraction. This means that in a dataset that has some long branches leading to the leaves, in addition to some other branches which are short, the long branches will attract each other and appear as a cluster in the tree, even if they are not closely related. This results from the fact that it is more favorable from a scoring point of view, to cluster dissimilar sequences together than to place them separately at the root. The phenomenon of long-branch attraction is known to affect in particular, maxi-mum parsimony algorithms although this type of artefact can present in various types of phylogenetic tree regardless of the algorithm used.

1.6.2
Rhodopsin β-Group

The β-group in the Rhodopsin family is not subdivided further into named sub branches. All the known ligands to the receptors in this cluster are peptides. The group includes the branch containing orexin-, neuropeptide FF-, tachykinin-, cholecystokinin-, prolactin-releasing hormone receptor and the neuropeptide Y receptors. As can be seen in Fig. 1.7, these receptors are relatively closely related as they place on the same branch in the tree. On the same branch, containing 18 receptors in total, the four orphan receptors GPR 72, GPR 73, GPR 73L1 and GPR 103 are also found. Continuing clockwise in Fig. 1.7, we find three orphan receptors, GPR, GPR 19 and GPR 154 that do not cluster with any other receptor but rather branch from the center of the tree. The next main cluster contains the neurotensin, motilin, ghrelin, and neuromedin receptors. This group also contain the orphan receptor GPR 39. Continuing further we find two orphan receptors GPR 75 and GPR 150 that do not clearly group with any other receptors. As can be seen the branches for these receptors are among the longest in the entire tree which shows that they are only distantly related to all the other receptors in the cluster. This is especially pronounced for GPR 75. The next group contains the gonadotropin releasing hormone, angiotensin, vasopressin, and oxytocin receptors. And finally the last group contains the endothelin, neuromedin B, gastrin releasing peptide, and the bombesin-like 3 receptors. Noteworthy in this group is that the neuropeptide Y receptors show a clearly heterogeneous phylogeny. In particular, the neuropeptide Y receptor 2 is not placed together with the other NPY receptors but rather with the tachykinin, prolactin-releasing hormone receptors and orphan receptors GPR 72, 73 and 73L1. It can also be seen that the neuropeptide Y receptor 2 has a higher amino acid identity to prolactin-releasing hormone receptor (36.5%) and GPR 72 (33.9%) than to the other NPY receptors (30.9, 34.2, and 30.9% to neuropeptide receptors 1, 4 and 5 respectively). The neuropeptide Y receptor 5 places equally distant to the neuropeptide Y 1 and 4 pair of receptors and cholecystokinin receptors. The reason for this is probably that neuropeptide Y receptor 5 has a long third extracellular loop due to an insertion, which is also present in the cholecystokinin receptor. The evolutionary history of this group of peptide binding receptors is likely to be much more complicated than it appears at first. For example, the prolactin-releasing hormone receptor is likely to share a recent evolutionary ancestor with the NPY receptors (Lagerstrom et al., unpublished data).

1.6.3
Rhodopsin γ-Group

This group comprises three main clusters termed the SOG, MCH, and Chemokine receptor clusters (Fig. 1.8).
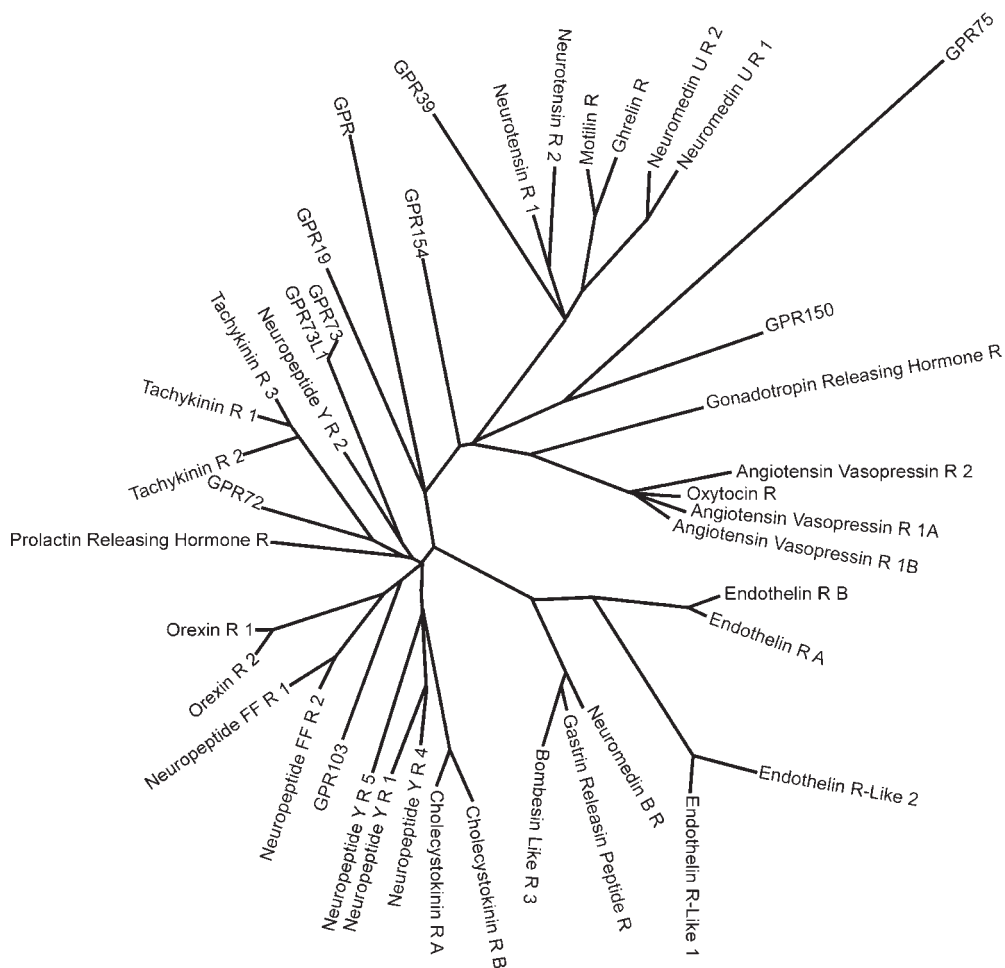
Fig. 1.7 Phylogenetic trees for each family and group of GPCRs. For further information see legend Fig. 1.1.

1.6.3.1 The SOG Receptor Cluster

This cluster of receptors contains the somatostatin, opioid, galanin, and neuro-peptide W receptors. It also contains the former orphan receptor GPR 54 which is now known to bind RF-amid. The known ligands to the receptor on this branch are thus all peptides but these ligands have, apart from their peptide nature, no structural similarities. The receptors in this cluster form four phylogenetic clusters (see Fig. 1.8), one containing only the five somatostatin receptors, one containing the two neuropeptide W receptors, one containing the four opioid receptors and finally one containing the three galanin receptors and GPR 54.

Fig. 1.8 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

1.6.3.2   **The Melanocyte Concentrating Hormone Receptor Cluster**
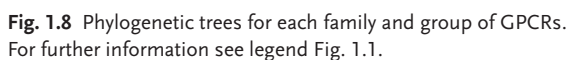
The two melanocyte concentrating hormone (MCH) receptors branch rather basally in the SOG cluster with relatively long branches. The branches connecting the two MCH receptor leafs are also relatively long which indicates that either these two receptors split quite early from a common ancestor or that they evolved

at a much faster rate than the other receptors in the SOG cluster. The ligand is MCH which is a cyclic neuropeptide of 19 amino acids that is involved in, among other things, the regulation of feeding behavior.

1.6.3.3 The Chemokine Receptor Cluster

This cluster is by far the largest in the Rhodopsin γ-group. The branch with most receptors contains the classical chemokine receptors of the C-C (12 receptors) and C-X-C types (eight receptors). These receptors are known to bind small immuno-modulating peptides. At the base of the chemokine branch is a relatively long branch containing the adrenomedulin receptor and one orphan chemokine-like receptor has split off. Continuing clockwise along the tree in Fig. 1.8 is a cluster containing the bradykinin and the two angiotensin receptors. The next branch contains three orphan receptors, the angiotensin-like receptor 1, GPR 15 and GPR 25. Further, the next branch contains the orphan receptor GPR 100 and the relaxin 3 receptor 1. Finally the Chemokine branch contains another large branch containing three complementary component receptors, three formyl peptide receptors, three orphan receptors and one chemoattractant receptor.

1.6.3.4 Other Rhodopsin γ-Receptors

Similar to the Rhodopsin α-group there are a number of mainly orphan receptors connected close to the center of the tree in a loose cluster with long branches. Using reasoning similar to that for Other Rhodopsin α-group receptors these should probably be considered to be more or less unrelated. The receptors in this "branch" are the two leukotriene B4 receptors and 10 orphan receptors namely, GPR 141, GPR 120, GPR 135, GPR 2037, GPR 146, GPR 35, chemokine-like receptor 2, GPR 139, GPR 31, GPR 14 and GPR 152.

1.6.4
The Rhodopsin δ-Group

This group (Fig. 1.9) has five main branches termed the MAS-related receptor cluster, the Glycoprotein receptor cluster, the Purinergic receptor cluster, the Co-agulation factor receptor cluster and the Olfactory receptor cluster (not shown in Fig. 1.9).

1.6.4.1 The MAS-related Receptor Cluster

This group contains 10 receptors with the MAS oncogene receptor and the MAS-related receptor branching most basally in the cluster. The four MAS-related GPCR D-G are separated by relatively long branches while the MAS related GPCR X1–4 are clearly highly similar. The MAS related GPCR X receptors have been shown to bind small peptide fragments originating from opioid peptides and these receptors are mainly expressed on small sensory neurons [34].
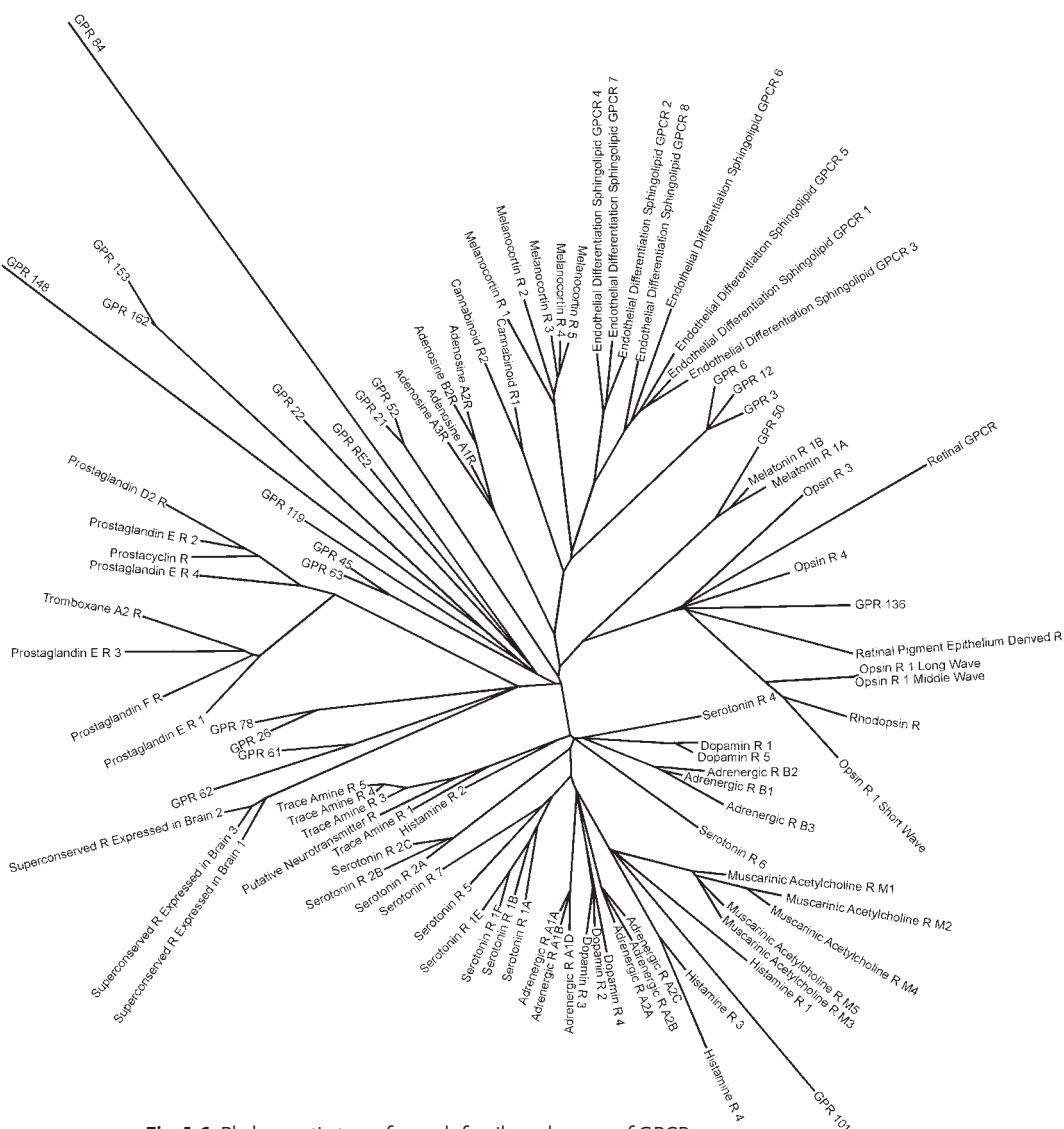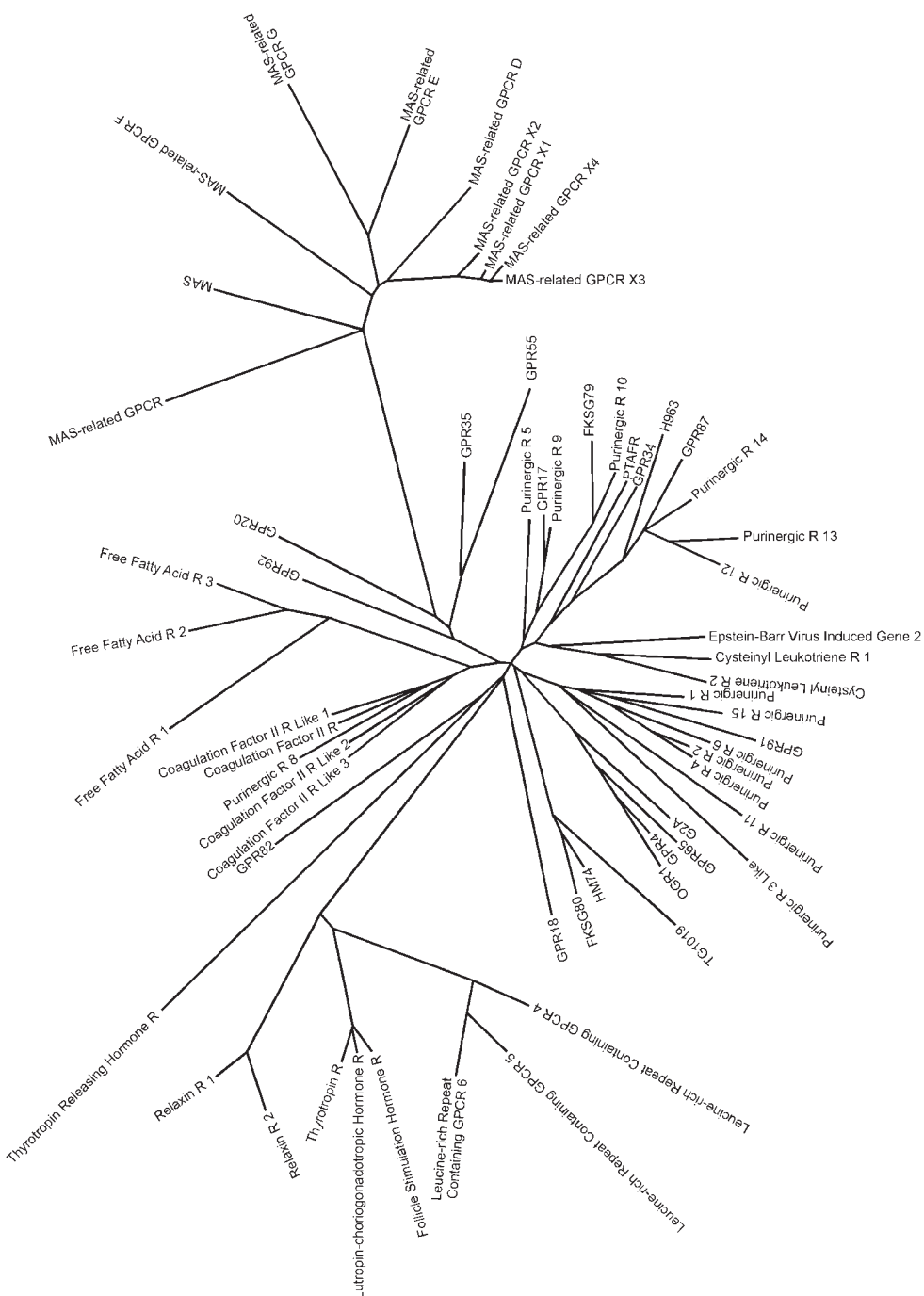
Fig. 1.9 Phylogenetic trees for each family and group of GPCRs.
For further information see legend Fig. 1.1.

1.6.4.2 **The Glycoprotein Receptor Cluster**

This cluster of receptors contains the classical glycoprotein hormone receptors, follicle stimulating hormone receptor, tyrotropin releasing hormone receptor and the leutropin-choriogonadotropin hormone receptor. It also contains the two relaxin receptors and three orphan receptors. The receptors in this cluster are different from all other Rhodopsin receptors because they have long N-termini which contain the ligand recognition domain, in common with most non-Rhodopsin GPCRs. This cluster also contains the two orphan receptors GPR 18 and GPR 82.

1.6.4.3 **The Coagulation Factor Receptor Cluster**

This cluster contains four receptors for coagulation factors, one purinergic receptor and a separate branch with the three free fatty acid binding receptors (previously known as GPR 40, 41 and 43) [35–37].

1.6.4.4 **The Purinergic Receptor Cluster**

This large branch consists of 13 nucleotide binding, or purinergic, receptors (P2Ys), the platelet activating factor receptor, the two cysteinyl leuktriene receptors and 14 orphan GPCRs and hence the known ligands for the receptors in this group are extracellular nucleotides, leukotrienes and thrombin. Considering the close phylogenetic relationship between the receptors with known ligands and the large number of orphan receptors it is likely that additional purinergic receptors will be discovered among the orphan GPCRs in the Purinergic cluster.

1.6.4.5 **The Olfactory Receptor Cluster**

The olfactory receptors clearly belong to the Rhodopsin γ-group using both phylogenetic and sequence clustering methods and our phylogenetic analysis indicates that the olfactory receptors form a stable phylogenetic cluster, which does not overlap with other groups of the rhodopsin family or with other families. Our previous searches in the human genome databases indicated that there could be over 460 olfactory receptors in the human genome that we consider likely to represent unique functional receptors [9]. A total of 347 putative human full-length odorant receptor genes have previously been identified and physically cloned [25]. It has also been suggested that there are over 900 olfactory receptor-like sequences in the human genome [2]. Further work is needed to identify the total number of olfactory receptors.

1.6.4.6 **Other Rhodopsin α-Receptors**

As is the case with the Rhodopsin α and γ groups, the Rhodopsin δ group contains four receptors that cannot be positioned phylogenetically on any branch of the tree. These are the orphan receptors GPR 35, GPR 55, GPR 20 and GPR 92.

1.7
Other GPCRs

There are currently 19 GPCRs that either cannot logically be placed into any of the main families (eight) or are likely to belong to the Rhodpsin family but cannot with certainty be placed into any of the four groups (11). Of these putative rhodopsin receptors six (V1RL1, V1RL2, V1RL4, FKSG46, FKSG83 and hGPCR23) are clearly related to each other and hence form a small "cluster". These are all related to vomeronasal receptors type 1, a family of GPCRs with over 100 members in for example the mouse and rat genomes. As humans do not appear to have a vomeronasal system for pheromone detection it is presently unclear whether these five human receptors have a physiological function, although they appear to be functional based on their amino acid sequence, or if they perform some other function in humans unrelated to pheromone detection. The other receptors putatively belonging to the Rhodopsin family are the Duffy receptor, known to be involved in immune system modulation and the four orphan receptors GPR 88, GPR 142, GPR 160 and hGPCR 19. Of the non-Rhodopsin receptors in the "Other" group two pairs of related receptors can be seen. TM7SF1 and C11ORF4 are clearly related to each other as are PERVAR1 and PERVAR2. Interestingly none of these receptors has any detectable sequence identity to any GPCR outside its pair and the same is true for the last four "Other receptors", IEDA, OA1, hGPCR29 and hGPCR43.

1.8
Future Perspective

There are currently 19 GPCRs that either can not be placed clearly into any of the main families (eight) or are likely to belong to the *Rhodopsin* family but can not with certainty be placed into any of the four groups (eleven). Much work has recently been carried out to organize the "functional" protein coding gene repertoire of GPCRs in the human genome. There is however considerable work remaining. Even though most of the genes have been identified, their exact genomic structure, including all alternatively spliced isoforms, needs to be worked on, in particular for those receptors that have recently been identified. Which genes are pseudogenes and which are not? Do many of the pseudogenes have a functional role even though they do not look like GPCRs with intact seven transmembrane regions? Our recent understanding of RNA genes has also emphasized the need to adopt a more "open minded" approach to what is functional and what is not. This is an important problem among those groups of receptors that are either rapidly expanding or shrinking in numbers in the human genome. Some groups such as the olfactory receptors clearly need more work while the exact domain composition of other groups such as the relatively newly established adhesion family also need more fine tuning. Progress is being made on the repertoire of GPCR in other genomes. In a recent article we created Hidden Markov Models

based on the different groups of human GPCRs and added several other models based on receptors not found in mammals [9, 38]. We searched the entire Genscan datasets from 13 species whose genomes are nearly completely sequenced. We reported over 5000 unique GPCRs that were divided into 15 main groups, and that the Rhodopsin family was subdivided into 13 subclasses. This chapter shows that all the five main families in the human genome arose prior to the split of nematodes from the chordate lineage and that several of the subgroups of the Rhodopsin family arose prior to the split of lineage which led to the vertebrates. We are currently working on more detailed analysis in each species with emphasis on the mouse, rat, and chicken genomes. This will provide a better understanding of the origin of the human GPCRs and the subgroup-specific changes that have occurred during the last 300 million years.

Table 1.1 All 19 receptors in the group "Other" were searched against a database, using BLASTP with a cut-off at 0.01, containing only all human GPCRs and the best non-self hit was recorded. The first column shows the overall best hit and the second the best hit when all receptors from the "Other" group of are excluded. E values from the BLASTP searches are presented in parentheses. A putative family assignment based on the hit in the second column is also presented.

Receptor name	Best human BLAST-hit	Best human BLAST-hit excluding "Other""	Putative main family
C11ORF4	TM7SF1 (3.0e-076)	NA	NA
DUFFY	F2RL2 (0.001)	F2RL2 (0.001)	Rhodopsin (δ)
FKSG46	V1RL2 (8.4e-082)	NA	Rhodopsin (δ)
FKSG83	FKSG46 (1.2e-031)	TRHR (2.1e-004)	Rhodopsin (δ)
GPR 142	GPR139 (1.4e-062)	GPR139 (1.4e-062)	Rhodopsin (γ)
GPR 160	TAR5 (1.2e-004)	TAR5 (1.2e-004)	Rhodopsin (α)
GPR 88	OPRL1 (1.7e-005)	OPRL1 (1.7e-005)	Rhodopsin (γ)
IEDA	NA	NA	NA
OA1	VIPR2 (3.0e-005)	GPR144 (0.006)	NA
PERVAR1	PERVAR2 (1e-139)	NA	NA
PERVAR2	PERVAR1 (3e-148)	NA	NA
TM7SF1	C11ORF4 (2.1e-096)	CCRL2 (0.001)	NA
V1RL1	FKSG46 (7.8e-071)	TRHR (1.5e-005)	Rhodopsin (δ)
V1RL2	V1RL4 (9.5e-089)	BRS3 (2.6e-006)	Rhodopsin (β)
V1RL4	V1RL2 (4.0e-089)	EDNRB (6.8e-004)	Rhodopsin (β)
hGPCR 19	EDG6 (6.2e-006)	EDG6 (6.2e-006)	Rhodopsin (α)
hGPCR 23	FKSG46 (4.0e-015)	NA	Rhodopsin (δ)
hGPCR 29	NA	NA	NA
hGPCR 43	NA	NA	NA

Table 1.2 The number of GPCRs in each of the main families. The table also shows the number of receptors in each family that recognize ligands of the types, peptide, biogenic amine, lipids and other. The last column designates the number of orphan receptors in each of the families.

Group	Number	Peptide ligand	Biogenic amine ligand	Lipid ligand	Other ligand	Orphan
Adhesion	33	0	0	0	1	32
Secretin	15	15	0	0	0	0
Frizzled	11	0	0	0	11	0
Taste type 2	26	0	0	0	1	25
Glutamate	22	0	10	0	4	8
Rhodopsin (α)	101	7	40	22	7	25
Rhodopsin (β)	42	32	0	0	0	10
Rhodopsin (γ)	72	54	0	0	2	18
Rhodopsin (δ)	63	17	0	6	14	26
Other	20	1	0	0	0	19
Total	407	126	50	28	40	163

Acknowledgments

References

1 Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., et al. Initial sequencing and analysis of the human genome. *Nature* **2001**; *409*: 860–921.

2 Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., et al. The sequence of the human genome. *Science* **2001**; *291*: 1304–1351.

3 Fredriksson, R., Schioth, H. B. The repertoire of G-protein coupled receptors in fully sequenced genomes. *Mol Pharmacol* **2005**; *67*: 1414–1425.

4 Attwood, T. K., Findlay, J. B. Fingerprinting G-protein-coupled receptors. *Protein Eng* **1994**; *7*: 195–203.

5 Kolakowski, L. F., Jr. GCRDb: a G-protein-coupled receptor database *Receptors Channels* **1994**; *2*: 1–7.

6 Josefsson, L. G. Evidence for kinship between diverse G-protein coupled receptors. *Gene* **1999**; *239*: 333–340.

7 Flower, D. R. Modelling G-protein-coupled receptors for drug design. *Biochim Biophys Acta* **1999**; *1422*: 207–234.

8 Bockaert, J., Pin, J. P. Molecular tinkering of G protein-coupled receptors:

an evolutionary success. *EMBO J* **1999**; *18*: 1723–1729.

9 FREDRIKSSON, R., LAGERSTROM, M. C., LUNDIN, L. G., SCHIOTH, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

10 FREDRIKSSON, R., LAGERSTROM, M. C., HOGLUND, P. J., SCHIOTH, H. B. Novel human G protein-coupled receptors with long N-terminals containing GPS domains and Ser/Thr-rich regions. *FEBS Lett* **2002**; *531*: 407–414.

11 FREDRIKSSON, R., GLORIAM, D. E., HOGLUND, P. J., LAGERSTROM, M. C., SCHIOTH, H. B. There exist at least 30 human G-protein-coupled receptors with long Ser/Thr-rich N-termini. *Biochem Biophys Res Commun* **2003**; *301*: 725–734.

12 BJARNADOTTIR, T. K., FREDRIKSSON, R., HOGLUND, P. J., GLORIAM, D. E., LAGERSTROM, M. C., SCHIOTH, H. B. The human and mouse repertoire of the adhesion family of G-protein-coupled receptors. *Genomics* **2004**; *84*: 23–33.

13 FREDRIKSSON, R., HOGLUND, P. J., GLORIAM, D. E., LAGERSTROM, M. C., SCHIOTH, H. B. Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. *FEBS Lett* **2003**; *554*: 381–388.

14 MCKNIGHT, A. J., GORDON, S. The EGF-TM7 family: unusual structures at the leukocyte surface. *J Leukoc Biol* **1998**; *63*: 271–280.

15 LIN, H. H., STACEY, M., HAMANN, J., GORDON, S., MCKNIGHT, A. J. Human EMR2, a novel EGF-TM7 molecule on chromosome 19p13.1, is closely related to CD97. *Genomics* **2000**; *67*: 188–200.

16 HARMAR, A. J. Family-B G-protein-coupled receptors. *Genome Biol* **2001**; *2*: REVIEWS3013.

17 EICHLER, W. CD97 isoform expression in leukocytes. *J Leukoc Biol* **2000**; *68*: 561–567.

18 CHEY, W. Y., CHANG, T. M. Secretin, 100 years later. *J Gastroenterol* **2003**; *38*: 1025–1035.

19 BLAIR HEDGES, S., KUMAR, S. Genomic clocks and evolutionary timescales. *Trends Genet* **2003**; *19*: 200–206.

20 SLUSARSKI, D. C., CORCES, V. G., MOON, R. T. Interaction of Wnt and a Frizzled homologue triggers G-protein-linked phosphatidylinositol signalling. *Nature* **1997**; *390*: 410–413.

21 KRISTIANSEN, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther* **2004**; *103*: 21–80.

22 CHANDRASHEKAR, J., MUELLER, K. L., HOON, M. A., ADLER, E., FENG, L., GUO, W., ZUKER, C. S., RYBA, N. J. T2Rs function as bitter taste receptors. *Cell* **2000**; *100*: 703–711.

23 BUFE, B., HOFMANN, T., KRAUTWURST, D., RAGUSE, J. D., AND MEYERHOF, W. The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides. *Nat Genet* **2002**; *32*: 397–401.

24 RYBA, N. J., TIRINDELLI, R. A new multigene family of putative pheromone receptors. *Neuron* **1997**; *19*: 371–379.

25 ZOZULYA, S., ECHEVERRI, F., NGUYEN, T. The human olfactory receptor repertoire. *Genome Biol* **2001**; *2*: RESEARCH0018.

26 PALCZEWSKI, K., KUMASAKA, T., HORI, T., BEHNKE, C. A., MOTOSHIMA, H., FOX, B. A., LE TRONG, I., TELLER, D. C., et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

27 LAGERSTROM, M. C., KLOVINS, J., FREDRIKSSON, R., FRIDMANIS, D., HAITINA, T., LING, M. K., BERGLUND, M. M., SCHIOTH, H. B. High affinity agonistic metal ion binding sites within the melanocortin 4 receptor illustrate conformational change of transmembrane region 3. *J Biol Chem* **2003**; *278*: 51521–51526.

28 BERGLUND, M., FREDRIKSSON, R., SALANECK, E., LARHAMMAR, D. Reciprocal mutations of neuropeptide Y receptor Y2 in human and chicken identity amino acids important for agonist binding. *FEBS Lett* **2002**; *518*: 5–9.

29 Haitina, T., Klovins, J., Andersson, J., Fredriksson, R., Lagerstrom, M. C., Larhammar, D., Larson, E. T., Schioth, H. B. Cloning, tissue distribution, pharmacology and three-dimensional modelling of melanocortin receptors 4 and 5 in rainbow trout suggest close evolutionary relationship of these subtypes. *Biochem J* **2004**; *380*: 475–486.

30 Baldwin, J. M. Structure and function of receptors coupled to G proteins. *Curr Opin Cell Biol* **1994**; *6*: 180–190.

31 Nathans, J., Thomas, D., Hogness, D. S. Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science* **1986**; *232*: 193–202.

32 Thomson, P. A., Wray, N. R., Thomson, A. M., Dunbar, D. R., Grassie, M. A., Condie, A., Walker, M. T., Smith, D. J., et al. Sex-specific association between bipolar affective disorder in women and GPR50, an X-linked orphan G protein-coupled receptor. *Mol Psychiatry* **2004**; *10*: 470–478.

33 Uhlenbrock, K., Gassenhuber, H., Kostenis, E. Sphingosine 1-phosphate is a ligand of the human gpr3, gpr6 and gpr12 family of constitutively active G protein-coupled receptors. *Cell Signal* **2002**; *14*: 941–953.

34 Lembo, P. M., Grazzini, E., Groblewski, T., O'Donnell, D., Roy, M. O., Zhang, J., Hoffert, C., Cao, J., Schmidt, R., Pelletier, M., Labarre, M., Gosselin, M., Fortin, Y., Banville, D., Shen, S. H., Strom, P., Payza, K., Dray, A., Walker, P., Ahmad, S. Proenkephalin A gene products activate a new family of sensory neuron-specific GPCRs. *Nat Neurosci* **2002**; *5*: 201–209.

35 Nilsson, N. E., Kotarsky, K., Owman, C., Olde, B. Identification of a free fatty acid receptor, FFA2R, expressed on leukocytes and activated by short-chain fatty acids. *Biochem Biophys Res Commun* **2003**; *303*: 1047–1052.

36 Kotarsky, K., Nilsson, N. E., Flodgren, E., Owman, C., Olde, B. A human cell surface receptor activated by free fatty acids and thiazolidinedione drugs. *Biochem Biophys Res Commun* **2003**; *301*: 406–410.

37 Brown, A. J., Goldsworthy, S. M., Barnes, A. A., Eilert, M. M., Tcheang, L., Daniels, D., Muir, A. I., Wigglesworth, M. J., et al. The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J Biol Chem* **2003**; *278*: 11312–11319.

38 Perez, D. M. From Plants to Man: The GPCR „Tree of Life" (Relates to article by Fredriksson, et al., FastForward 1 February 2005). *Mol Pharmacol* **2005**; *67*: 1383–1384.

2

Why G Protein-coupled Receptors Databases are Needed

Jacques Haiech, Jean-Luc Galzi, Marie-Claude Kilhoffer, Marcel Hibert,
and Didier Rognan

2.1
Introduction

G protein-coupled receptors probably comprise one of the largest families of human receptors. This class of receptor represents the therapeutic targets of more than 30 % of the drugs now on the market [1].

G protein-coupled proteins are composed of seven transmembrane domains associated with NH_2 and COOH terminus domains. Numerous data have been collected in different databases and numerous lists of GPCRs are circulating on the web. The two main databases are the IUPHAR receptor database (http://www.iuphar-db.org/ iuphar-rd) and the GPCR-DB; (http://www.gpcr.org/7tm/). However, most of the data are not complete and moreover, these databases aim to provide a general overview and therefore cannot fulfill the specific requests of all types of user. For instance, the aforementioned databases are not ligand oriented, therefore in order to provide answers to ligand-associated queries, a ligand-oriented database has been compiled (http://gdds.pharm.kyoto-u.ac.jp/services/glida/).

A database uses a data model that is derived from a biological model. In the following section, we will attempt to define a typology of the different GPCR databases. In order to fulfill the needs of a set of specific users, data representation is needed. A data model defines the manner in which the data is represented. Our aim in this chapter is to illustrate the links between data models and data representation in the GPCR field.

2.2
A Non-exhaustive List of the GPCR Data Models

One of the main problems that we face is the establishment of a unified nomenclature of potential human GPCR genes. As a starting point we have used data

gathered from Uniprot and GPCRDB [2] and have focused on the endo-GPCR genes, which means that the olfactory, taste and vomeronasal receptors have been omitted.

As far as possible we have used the HUGO [3] and IUPHAR nomenclatures to name the genes. The locus ID appears to be the most practicable identification number.

Some databases use the Swissprot name and the Swissprot identification number. We have established a correlation table between the different GPCR nomenclatures (see Appendix 2.1). This list will be used throughout this chapter to illustrate our points.

2.3
Using the Central Dogma of Biology

The central dogma of molecular biology is illustrated in Fig. 2.1 together with a simple data model which has been based on it. Most of the generalist databases in biology are compiled using this data model; the main databases are the NCBI database [4, 5], the UCSC [6] and the EMBL ensemble database. This basic data model may be enriched by adding sub-models or gathering databases from different species.

In this data model, the main difficulty is to define the concept of genes. We will use the following definition: a gene is a DNA segment that contains all the information required to generate the different known transcripts.

The HUGO definition (a gene is a DNA segment that contributes to phenotype/function) is broader but less operational (http://www.gene.ucl.ac.uk/nomenclature/guidelines.html#Introduction).

Numerous data representations have been developed by different laboratories. Cards have been largely used (genecards, famDbtools [7, 8]) but maps have also been employed as exemplified by the browser of the UCSC or the Ensembl [9] databases or mapview from NCBI (http://www.ncbi.nlm.nih.gov/About/outreach/gettingstarted/mapviewer/). Figure 2.2 illustrates such data representation.

Using such a data model, information such as the distribution of the GPCR gene on the different chromosomes (numbers and positions) may be obtained.

Figure 2.3 illustrates the distribution of the human GPCR genes on the different chromosomes.

It is noteworthy that chromosome 21 does not possess any GPCR genes.

This distribution is not random on any given chromosome and GPCR genes are sometimes clustered together on a chromosome. A group of GPCR genes is considered to be clustered if the distance between two consecutive genes is less that 200,000 base pairs and if the cluster contains more than three genes. Table 2.1 summarizes the GPCR clusters in the human genome. The gene name follows our internal nomenclature (see Appendix 2.1 for the synonyms). It is noticeable that the gene clusters are mainly composed of GPCRs from the same functional cluster.
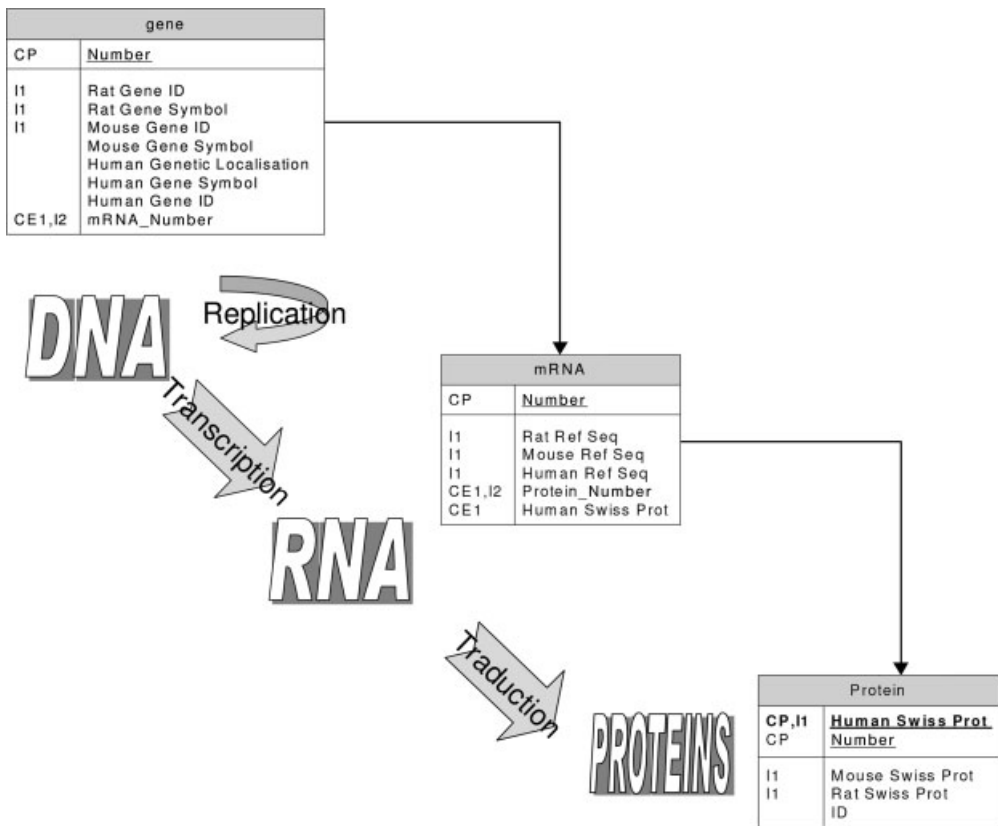
Fig. 2.1 Schematic data model using the central dogma of molecular biology.

It can be seen that the clusters occur on chromosomes 3, 6 and 19 and concern only the genes belonging to the rhodopsin or adhesion class. The first cluster is composed of eight genes belonging to the chemokine family. Several other observations may be made from these types of database, such as the expression profiles using the GEO database [10] (http://www.ncbi.nlm.nih.gov/projects/geo/), the SNP map (http://www.ncbi.nlm.nih.gov/projects/SNP/) or the distribution of the number of non-coding or coding exons in this family by comparing the mRNA sequences or the protein sequences with the genomic sequences. For instance, various tests can be applied to a section of DNA to elucidate specific areas such as the segment of DNA is an object on which we applied different methods to find specific patterns such as promoters, enhacers, exon–intron junctions, etc. (http://www.expasy.org/tools/).

We will use the chemokine cluster to present and illustrate some other database typologies [11–13].

Fig. 2.2a Data representation of a human gene using a card representation
(FamDBTool) [8].

2.4
Using the Tree of Life

The existence of a gene cluster may be due to a recent duplication of genes or to
the fact that the proximity of these genes proved advantageous during the evolu-
tion. Insight into the reasons for the existence of clusters can be gained by follow-
ing the behavior of such clusters during the evolution.

In this case, a database needs to be compiled based on a data model which takes
into account the classification and evolution of the species; such databases are ex-
emplified by ACNUC, [14] and Taxonomy browser, NCBI. Orthologs (homologous
genes between species) and paralogs (homologous genes in a given organism)
need to be grouped together and tools have been developed that facilitate this pro-
cedure (Blinks at NCBI) leading to the compilation of databases which group
orthologous or paralogous genes (http://doop.abc.hu/)[15].

Figure 2.4 illustrates a simple data model and the use of such databases to com-
pare the CCR4 – a chemokine receptor – protein sequences from different species.
The result of this comparison is presented as a phylogenic tree (Fig. 2.4).
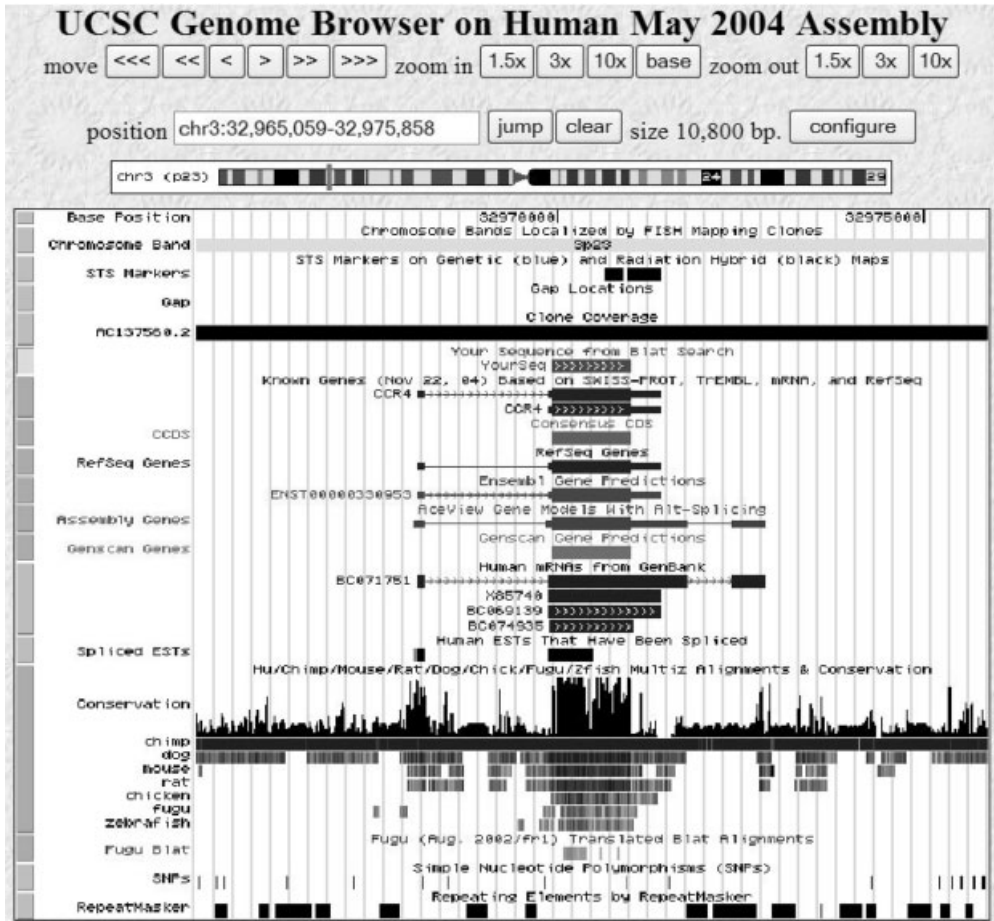
Fig. 2.2 b Data representation using a mapview representation.

This phylogenic tree illustrates the point that the chemokine receptor CCR4 is only present in bony vertebrates [16–19].

This chemokine receptor is part of a chemokine receptor gene cluster localized on human chromosome 3. A simple database can be used to collect the gene information regarding this cluster from several species (Table 2.2).

From this human gene cluster we have devised a phylogenic tree by comparing the protein sequences (Fig. 2.5).

This tree shows that the paralogous group of genes (CCRL2, CCR1, CCR3, CCR2 and CCR5) derives from a common ancestor probably by four duplication events. By comparing the orthologous genes of this group in the chicken, we observe that it exists a unique gene orthologous both to CCR1 and to CCR3 and a unique gene orthologous both to CCR2 and to CCR5. This suggests that the two last duplications took place after the separation of birds and mammals [16].
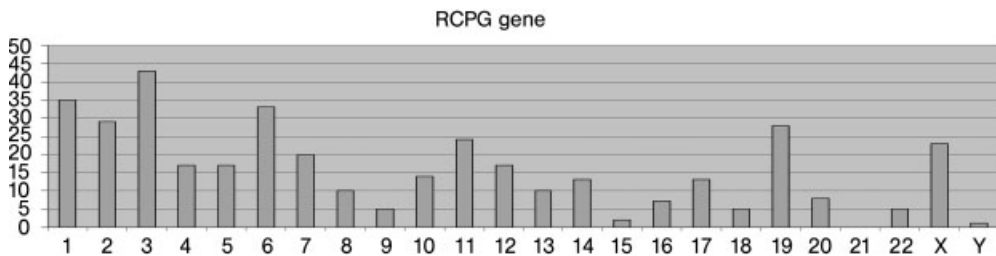
Fig. 2.3 The number of GPCR genes is indicated for each human chromosome. It is noteworthy that chromosome 21 does not possess any GPCR gene.

Table 2.1 Some human GPCR genes are clustered in the genome. This table gathered together the existing human GPCR chromosomal clusters composed of more than three genes.

Gene name	Class	Functional cluster	LocusID	Contig
CCR9	Rhodopsin	Chemokines	10803	chr3:45902023-45920671
CXCR6	Rhodopsin	Chemokines	10663	chr3:45958977-45965849
XCR1	Rhodopsin	Chemokines	2829	chr3:46036442-46039443
CCR1	Rhodopsin	Chemokines	1230	chr3:46217732-46225819
CCR3	Rhodopsin	Chemokines	1232	chr3:46257692-46284166
CCR2	Rhodopsin	Chemokines	1231	chr3:46369644-46376913
CCR5	Rhodopsin	Chemokines	1234	chr3:46385637-46393695
CCRL2	Rhodopsin	Chemokines	9034	chr3:46422754-46427017
GPR171	Rhodopsin	Purines	29909	chr3:152397318-152404677
GPR105	Rhodopsin	Purines	9934	chr3:152413789-152414802
GPR87	Rhodopsin	Purines	53836	chr3:152493595-152518334
GPR86	Rhodopsin	Purines	53829	chr3:152525794-152531033
P2RY12	Rhodopsin	Purines	64805	chr3:152537074-152586242
TRAR3	Rhodopsin	Amines	134860	chr6:132900120-132903168
TRAR5	Rhodopsin	Amines	83551	chr6:132914525-132917553
TRAR4	Rhodopsin	Amines	319100	chr6:132932154-132935191
PNR	Rhodopsin	Amines	9038	chr6:132950505-132953518
GPR57	Rhodopsin	Amines	9288	chr6:132971082-132972108
GPR58	Rhodopsin	Amines	9287	chr6:132978982-132981902
TRAR1	Rhodopsin	Amines	134864	chr6:133006816-133009835
CD97	Adhesion	Adhesion	976	chr19:14352266-14381533
PTGER1	Rhodopsin	Prostanoid	5731	chr19:14443279-14448174
EMR3	Adhesion	Adhesion	84658	chr19:14590052-14647810
EMR2	Adhesion	Adhesion	30817	chr19:14703205-14751353
GPR40	Rhodopsin	Purines	2864	chr19:40533295-40536197
GPR41	Rhodopsin	Purines	2865	chr19:40540342-40544227
GPR42	Rhodopsin	Purines	2866	chr19:40553102-40556142
GPR43	Rhodopsin	Purines	2867	chr19:40631457-40634449
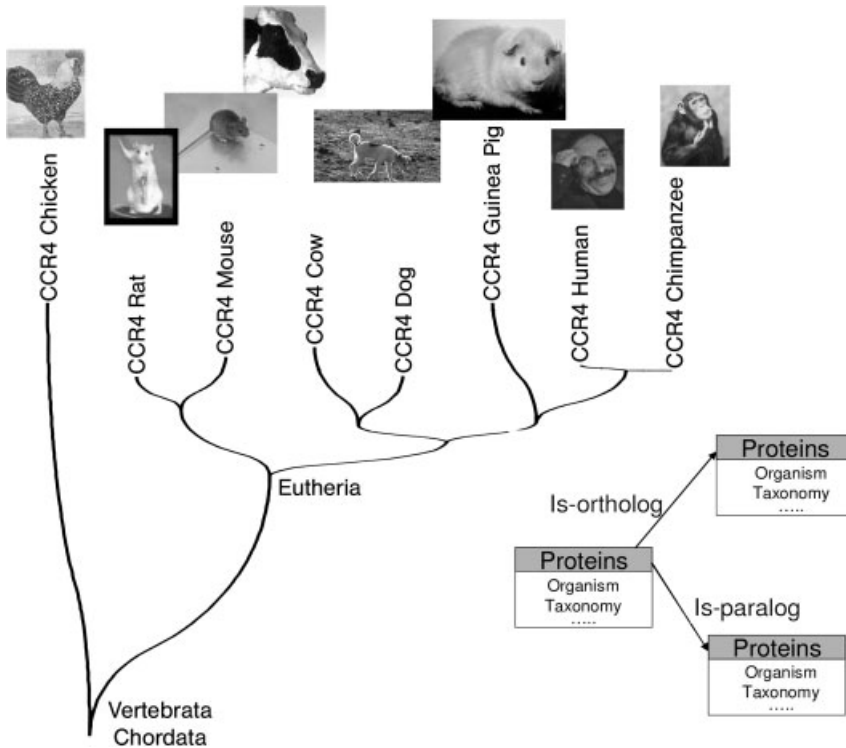
Fig. 2.4 Phylogenic tree derived after alignment of CCR4 orthologs by using the Mega3 software and the neighbor-joining method [20].
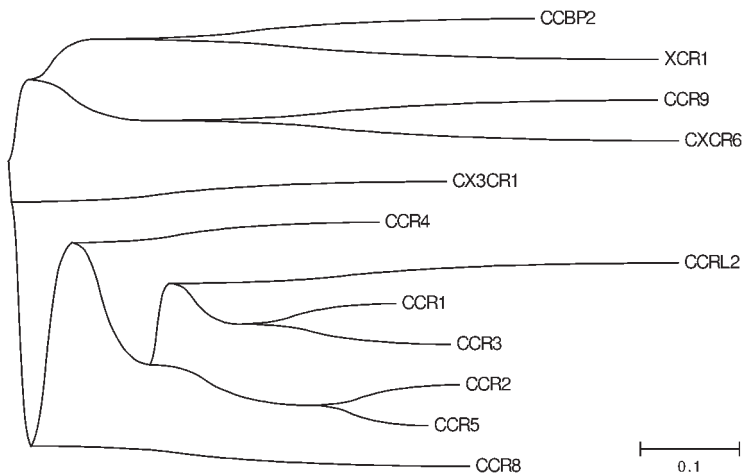
Fig. 2.5 Phylogenic tree of the human chemokine receptor gene cluster on chromosome 3.
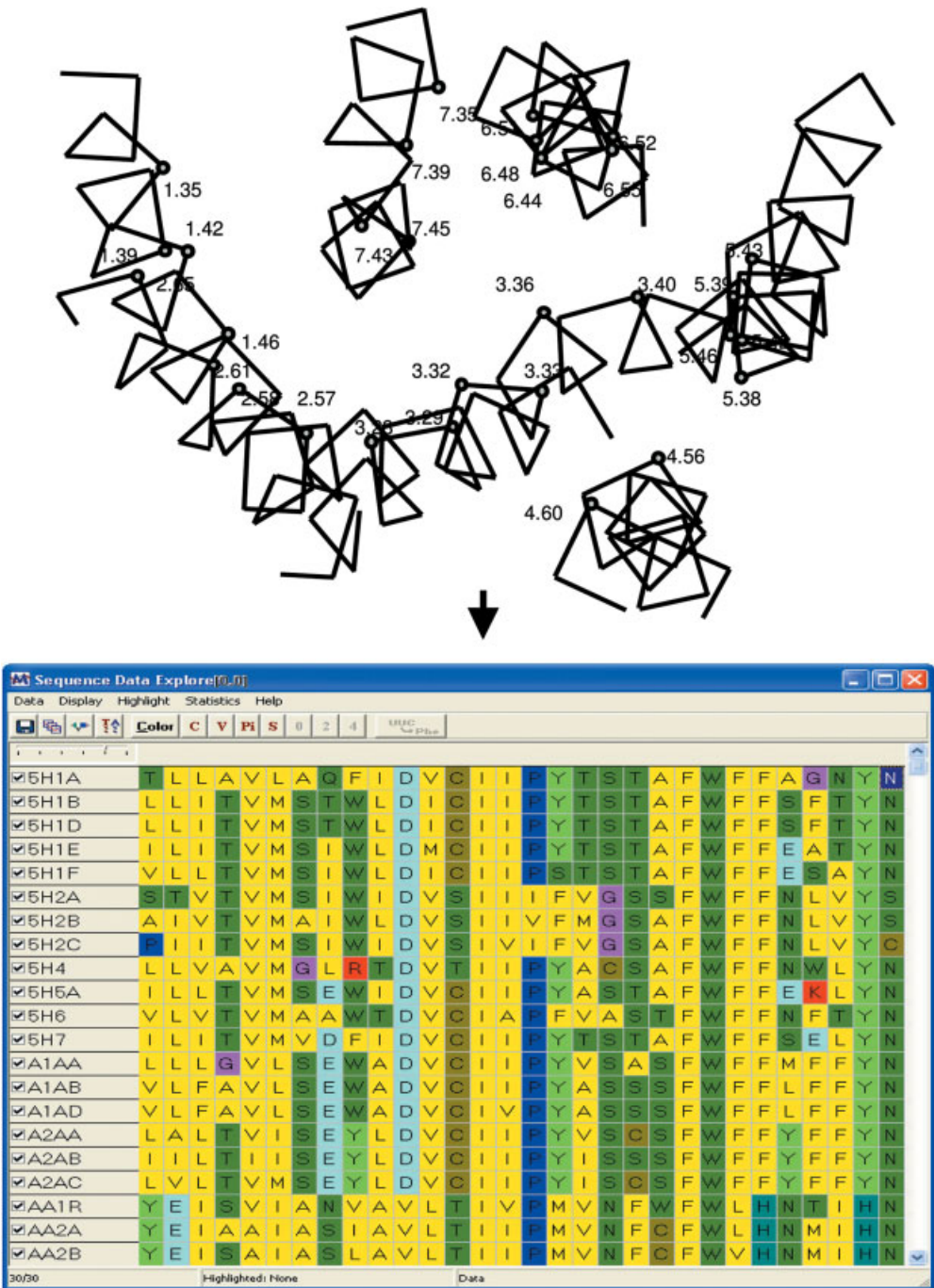
Fig. 2.6 (legend see page 35)
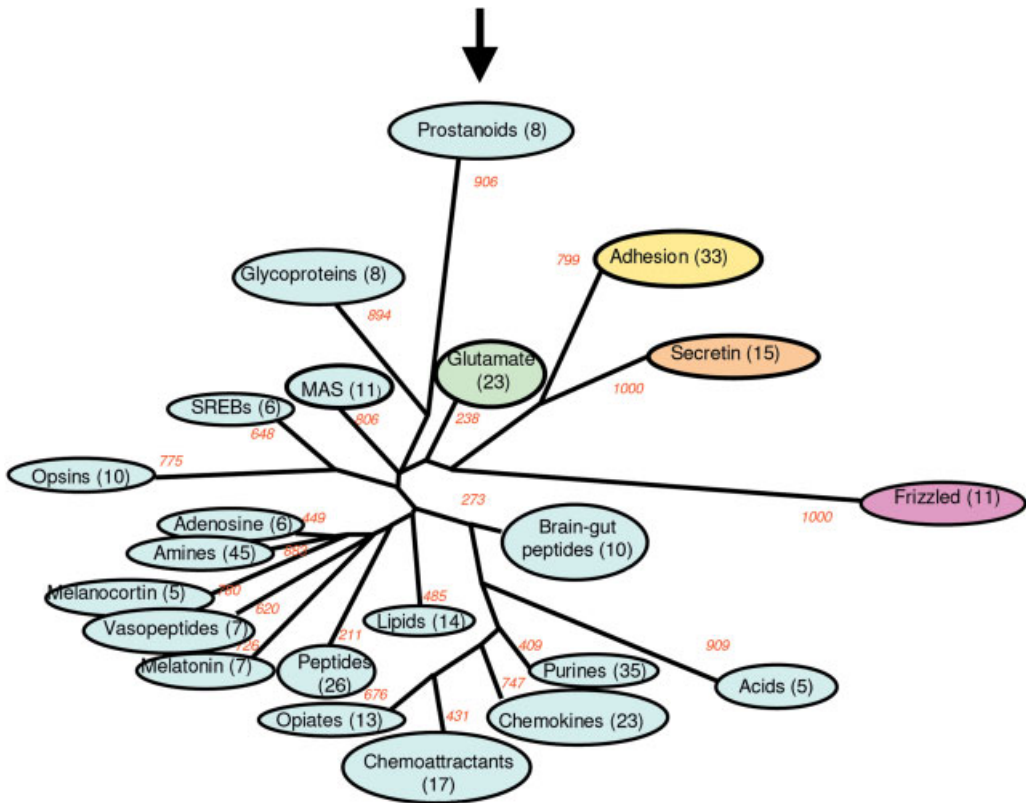
Fig. 2.6 Protocol to generate a TM cavity-driven phylogenetic tree: (1) selection of 30 critical positions, (2) definition of ungapped sequences comprising the 7-TM cavity, (3) TM cavity-derived phylogenetic tree for 372 human GPCRs. The consensus tree was derived from 1000 trees. Numbers in parentheses indicate the number of entries in each cluster. Numbers in italic represent bootstrap values to assess the statistical significance of the tree. Receptors classified as singletons (see text) are not displayed here for sake of clarity. Glutamate, Rhodospin, Adhesion, Frizzled and Secretin subfamilies are colored in green, cyan, yellow, pink and orange, respectively.

2.5
Using a Chemogenomic Approach

Assuming that similar GPCRs recognize similar ligands, an accurate annotation of all GPCR entries should take into account the similarities/differences at their ligand binding cavity. As most small molecular-weight ligands probably bind to the 7-TM core, all human GPCR entries have been annotated using a chemogenomic procedure [21] which fingerprints their 7-TM binding cavity. Thirty positions lining the retinal binding site in bovine rhodopsin were extracted from all entries and concatenated with ungapped sequences leading to the derivation of a phylogenetic tree using the UPGMA clustering method (Fig. 2.6).

Table 2.2 Localization of orthologs of the human chemokine gene cluster (in bold) and surrounding genes from mouse, rat, dog and chicken.

Name	LocusID	Human	Mouse	Rat	Dog	Chicken
GRM7	2917	chr3:6876927-7759217	chr6:111092457-112078094	chr4:146576419-147516066	chr20:13079459-13906450	chr12:18787324-19010925
OXTR	5021	chr3:8766094-8787300	chr6:114237807-114250527	chr4:148557930-148572638	chr20:12206658-12225278	chrUn:50576008-50576835
HRH1	3269	chr3:11268400-11280415	chr6:115004925-115006388	chr4:150675080-150678540	chr20:10008450-10009910	chr12:4432075-4433424
CCR4	1233	chr3:32967070-32972407	chr9:114418057-114424978	chr8:118901603-118904685	chr23:6851095-6852174	chr2:43489815-43503008
CX3CR1	1524	chr3:39278991-39297526	chr9:120034271-120055277	chr8:125052367-125068072	chr23:11945747-11946721	chr2:43488566-43489627
CCR8	1237	chr3:39345219-39352077	chr9:120078801-120081900	chr8:125095194-125096252	chr23:12025253-12026317	chr2:43465509-43467633
VIPR1	7433	chr3:42518121-42555064	chr9:121644916-121673565	chr8:126676972-126687362	chr23:14646497-14677366	chr2:1323205-1389069
CCBP2	1238	chr3:42880999-42882150	chr9:121901943-121916543	chr8:127045470-127051046	chr23:14972035-14973186	chr2:41805258-41805458
CCR9	10803	chr3:45902023-45920671	chr9:123795694-123796785	chr8:128546680-128562719	chr20:45462676-45486034	chr2:41880560-41881705
CXCR6	10663	chr3:45958977-45965849	chr9:123822395-123829572	chr8:128588301-128591353	chr20:45420818-45421840	chr2:41411921-41412100
XCR1	2829	chr3:46036442-46039443	chr9:123868239-123880049	chr8:128634640-128637578	chr20:45332174-45368102	chr2:41829361-41832763
CCR1	1230	chr3:46217732-46225819	chr9:123976524-123984607	chr8:128712826-128715893	chr20:45280484-45281548	chr2:41785007-41785993
CCR3	1232	chr3:46257692-46284166	chr9:124036503-124047343	chr8:128778680-128790067	chr20:45238497-45239573	chr2:41785010-41785942
CCR2	1231	chr3:46369644-46376913	chr9:124116692-124125059	chr8:128911521-128914642	chr20:45162939-45164054	chr2:41768162-41768977
CCR5	1234	chr3:46385637-46393695	chr9:124136440-124144048	chr8:128925894-128931948	chr20:45150588-45151643	chr2:41768030-41768989
CCRL2	9034	chr3:46422754-46427017	chr9:111058424-111063340	chr8:115470784-115473653	chr20:45125835-45133116	chr2:41785634-41785870
PTHR1	5745	chr3:46893240-46921291	chr9:110721018-110743628	chr8:115118119-115140954	chr20:44760449-44784424	chr2:3142172-3194416
CELSR3	1951	chr3:48647901-48676352	chr9:108830068-108858022	chr8:113901338-113944405	chr20:43312804-43338704	chr12:9074162-9095981
GRM2	2912	chr3:51717038-51728168	chr9:106664639-106673906	chr8:111858307-111867796	chr20:40765732-40774363	chr12:2573556-2577561
GPR62	118442	chr3:51963437-51967559	chr9:106481837-106484901	chr8:111626964-111630010	chr20:40609098-40611061	chrUn:10812551-10813489
GPR27	2850	chr3:71884891-71888018	chr6:100073278-100076417	chr4:134785375-134788508	chr20:40609833 -40610897	chr12:16222889-16223740
HTR1F	3355	chr3:88121984-88124690	chr16:65221879-65222976	chr11:2259587-2262687	chr31:3253398-3254495	chr1:86471817-86472905
GPR15	2838	chr3:99732568-99735650	chr16:58975816-58978865	chr11:42766517-42769596	chr33:8449590-8450669	chr1:77277131-77278111
GPR128	84873	chr3:101810135-101897955	chr16:56898553-56972217	chr11:44811741-44875874	chr33:10233108-10275484	chr1:79041018-79041158
DRD3	1814	chr3:115329247-115381446	chr16:43637739-43702417	chr11:58503490-58579178	chr33:21442151-21476476	chr1:77072338-77082143
GPR156	165829	chr3:121367569-121446635	chr16:37770496-37861308	chr11:64563569-64630360	chr33:26794550-26845753	chr1:74069423-74074912
CASR	846	chr3:123384220-123489032	chr16:36340851-36411288	chr11:66140758-66211881	chr33:28477902-28504357	chr1:69306745-69358367
RHO	6010	chr3:130729180-130737885	chr6:116317894-116326796	chr4:152301629-152308791	chr20:8476639-8481890	chr12:19431883-19433361

Twenty-two clusters could be unambiguously detected from the present analysis of 30 amino acid positions (Fig. 2.6). Of 372 entries 34 could not be assigned to one of the existing 22 clusters and are considered as singletons. The tree presented herein is very similar to the most complete phylogenetic tree (GRAFS classification) known to date [22], although the latter has been obtained from full TM sequences. In both classifications, GPCRs of the Frizzled, Glutamate, Secretin and Adhesion families cluster in well-separated groups whereas the large Rhodopsin family can be classified into 18 different clusters. Remarkably, all known GPCR subfamilies (e. g. receptors for biogenic amines, purines, and chemokines) are reproduced with high statistical significance using bootstrap. The five main families (Glutamate, Rhodopsin, Adhesion, Frizzled, Secretin) reported in the GRAFS classification are recovered with no overlaps between the corresponding clusters with the single exception of Q9GZN0 (GPR88), a rhodopsin-like GPCR clustered with class III GPCRs. Interestingly, receptors for which the orthosteric binding site is not located in the TM domain (Adhesion, Secretin and Glutamate families) are nevertheless grouped into homogeneous clusters. Relating cluster members to precise molecular features is here greatly facilitated by the analysis of a small subset of amino acids. For each of the 22 clusters, there is often a clear relationship between known ligand chemotypes (e. g. amines, carboxylic acids, phosphates, peptides, eicosanoids, and lipids) and the cognate TM cavities. For example, receptors for bulky ligands (e. g. phospholipids, prostanoids) have a TM cavity significantly larger than those among receptors for smaller compounds (e. g. biogenic amines, nucleotides). Receptors for charged ligands (cationic amines, phosphates, mono and di-carboxylic acids) always present among the 30 critical residues one or more conserved amino acids exhibiting the opposite charge (e. g. Asp3.32 for biogenic amines; Asp4.60/Glu7.39 for chemokines; Arg3.29/Lys6.55/Arg7.35 for nucleotides).

Our clustering approach implies two assumptions: (a) the overall fold of the 7-TM domain around the binding cavity has been conserved along evolution; (b) critical hotspots spread over the 7-TM domain repeatedly account for ligand binding. Although solid biostructural data for the three most important GPCR classes (class I, II, III) are missing, numerous experimental data do provide evidence in favor of strong similarities among many GPCRs: (a) residues known to affect small molecular-weight ligand binding to unrelated GPCRs are mostly spread among the selected 30 residues suggesting a common architecture of the TM pocket, (b) many known ligands are promiscuous for even unrelated GPCRs and are usually anchored through so-called privileged structures to common subpockets of different GPCRs [23]. Of course, we are aware that class II and class III GPCRs exhibit an additional orthosteric site located outside the 7-TM bundle. Therefore, conclusions drawn herein only apply to the 7-TM binding site.

2.6
Conclusion

Although numerous tools have been built to integrate the various biological data, there is no standard information system management in biology [24–30] that will allow the comparison and evaluation of different sources of information and the quality of the biological data produced by the different genomic centers.

Several web services deal with the GPCR family (for instance, [31, 32]) but do not provide a unified platform that will allow the integration of various data regarding this gene family. GPCR-DB [2] is the most successful in meeting the requirements of this type of database but still does not integrate data, for example, from ligand binding studies (agonists, antagonists and allosteric modulators) or information concerning interaction with others partners [33, 34]. Some integrative approaches have tried to compare sequence analysis and 3D-structure modeling [35]. We may expect that in the near future, a better link between the biological questions and the structure of the customized and specific databases will allow biologists to use the data that they are now producing in a more in-depth manner and at an accelerated pace due to the new high throughput technologies.

References

1 Drews J. Strategic trends in the drug industry. *Drug Discov Today* 2003; *8*: 411–420.

2 Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G. GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* 2003; *31*: 294–297.

3 Wain HM, Lush M, Ducluzeau F, Povey S. Genew: the human gene nomenclature database. *Nucleic Acids Res* 2002; *30*: 169–171.

4 Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2005; *33*: D54–D58.

5 Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Church DM, DiCuccio M, Edgar R, Federhen S, Helmberg W, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2005; *33*: D39–D45.

6 Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: The human genome browser at UCSC. *Genome Res* 2002; *12*: 996–1006.

7 Chalifa-Caspi V, Yanai I, Ophir R, Rosen N, Shmoish M, Benjamin-Rodrig H, Shklar M, Stein TI, Shmueli O, Safran M, et al. GeneAnnot: comprehensive two-way linking between oligonucleotide array probesets and GeneCards genes. *Bioinformatics* 2004; *20*: 1457–1458.

8 Haiech J, Moulhaye SB, Kilhoffer MC. The EF-Handome: combining comparative genomic study using FamDB-tool, a new bioinformatics tool, and the network of expertise of the European Calcium Society. *Biochim Biophys Acta* 2004; *1742*: 179–183.

9 Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, et al. Ensembl 2005. *Nucleic Acids Res* 2005; *33*: D447–D453.

10 Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R. NCBI GEO: mining millions

of expression profiles – database and tools. *Nucleic Acids Res* **2005**; *33*: D562–D566.

11 Comerford I, Nibbs RJ. Post-translational control of chemokines: a role for decoy receptors? *Immunol Lett* **2005**; *96*: 163–174.

12 Ribeiro S, Horuk R. The clinical potential of chemokine receptor antagonists. *Pharmacol Ther* **2005**; *107*: 44–58.

13 Schaerli P, Moser B. Chemokines: control of primary and memory T-cell traffic. *Immunol Res* **2005**; *31*:57–74.

14 Gouy M, Milleret F, Mugnier C, Jacobzone M, Gautier C. ACNUC: a nucleic acid sequence data base and analysis system. *Nucleic Acids Res* **1984**; *12*: 121–127.

15 Remm M, Storm CE, Sonnhammer EL. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J Mol Biol* **2001**; *314*: 1041–1052.

16 Wang J, Adelson DL, Yilmaz A, Sze SH, Jin Y, Zhu JJ. Genomic organization, annotation, and ligand-receptor inferences of chicken chemokines and chemokine receptor genes based on comparative genomics. *BMC Genomics* **2005**; *6*: 45.

17 Lio P, Vannucci M. Investigating the evolution and structure of chemokine receptors. *Gene* **2003**; *317*: 29–37.

18 DeVries ME, Cao H, Wang J, Xu L, Kelvin AA, Ran L, Chau LA, Madrenas J, Hegele RA, Kelvin DJ. Genomic organization and evolution of the CX3CR1/CCR8 chemokine receptor locus. *J Biol Chem* **2003**; *278*: 11985–11994.

19 Shields DC. Gene conversion among chemokine receptors. *Gene* **2000**; *246*: 239–245.

20 Kumar S, Tamurak Nei M. Mega 3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics* **2004**; *5*: 150–163.

21 Surgand JS, Rodrigo J., Kellenberger E, Rognan D. A chemogenomic analysis of the transmembrane binding cavity of human G protein-coupled receptors. *Proteins* **2005** (submitted).

22 Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

23 Bondensgaard K, Ankersen M, Thogersen H, Hansen BS, Wulff BS, Bywater RP. Recognition of privileged structures by G-protein coupled receptors. *J Med Chem* **2004**; *47*: 888–899.

24 Shah SP, Huang Y, Xu T, Yuen MM, Ling J, Ouellette BF. Atlas – a data warehouse for integrative bioinformatics. *BMC Bioinformatics* **2005**; *6*: 34.

25 Chabalier J, Capponi C, Quentin Y, Fichant G. ISYMOD: a knowledge warehouse for the identification, assembly and analysis of bacterial integrated systems. *Bioinformatics* **2005**; *i*:1 246–1256.

26 Barrera J, Cesar RM, Jr., Ferreira JE, Gubitoso MD. An environment for knowledge discovery in biology. *Comput Biol Med* **2004**; *34*: 427–447.

27 Liang F, Matrubutham U, Parvizi B, Yen J, Duan D, Mirchandani J, Hashima S, Nguyen U, Ubil E, Loewenheim J, et al. ORFDB: an information resource linking scientific content to a high-quality Open Reading Frame (ORF) collection. *Nucleic Acids Res* **2004**; *32*: D595–D599.

28 Helfrich JP. Raw data to knowledge warehouse in proteomic-based drug discovery: a scientific data management issue. *Biotechniques* **2002**; Suppl: 48–50, 52–43.

29 Wu OP, Seow KT, Wong L, Chung SY, Subbiah S. From sequence to structure to literature: the protocol approach to bioinformation. *Pac Symp Biocomput* **1998**; 747–758.

30 Eckman BA, Aaronson JS, Borkowski JA, Bailey WJ, Elliston KO, Williamson AR, Blevins RA. The Merck Gene Index browser: an extensible data integration system for gene finding, gene characterization and EST data mining. *Bioinformatics* **1998**; *14*: 2–13.

31 Papasaikas PK, Bagos PG, Litou ZI, Promponas VJ, Hamodrakas SJ. PRED-GPCR: GPCR recognition and family classification server. *Nucleic Acids Res* **2004**; *32*: W380–W382.

32 Peitsch MC, Herzyk P, Wells TN, Hubbard RE. Automated modelling of the transmembrane region of G-protein coupled receptor by Swiss-model. *Receptors Channels* **1996**; *4*: 161–164.

33 Sgourakis NG, Bagos PG, Papasaikas PK, Hamodrakas SJ. A method for the prediction of GPCRs coupling specificity to G-proteins using refined profile Hidden Markov Models. *BMC Bioinformatics* **2005**; *6*: 104.

34 Elefsinioti AL, Bagos PG, Spyropoulos IC, Hamodrakas SJ. A database for G proteins and their interaction with GPCRs. *BMC Bioinformatics* **2004**; *5*: 208.

35 Flower DR, Attwood TK. Integrative bioinformatics for functional genome annotation: trawling for G protein-coupled receptors. *Semin Cell Dev Biol* **2004**; *15*: 693–701.

Appendix 2.1

GPCR human gene using our internal nomenclature (FamDBtool), the Rognan nomenclature (Rognan-DB), the HUGO nomenclature and the IUPHAR nomenclature. To each gene is associated a locus ID. Class and cluster as determined by Rognan et al. are indicated. The table is subdivided according to the classes.

FamDBtool	UniProt	HUGO	Official IUPHAR nomenclature	Provisional name IUPHAR	Class	Cluster	Locus ID
ADORA1	AA1R	ADORA1	A1		Rhodopsin	Adenosine	134
ADORA2A	AA2AR	ADORA2A	A2A		Rhodopsin	Adenosine	135
ADORA2B	AA2BR	ADORA2B	A2B		Rhodopsin	Adenosine	136
ADORA3	AA3R	ADORA3	A3		Rhodopsin	Adenosine	140
GNRHR	GNRHR	GNRHR			Rhodopsin	Adenosine	2798
GNRHR2	GNRR2	GNRHR2			Rhodopsin	Adenosine	114814
GPR148		GPR148			Rhodopsin	Amines	344561
ADRA1A	ADA1	ADRA1A	a1A		Rhodopsin	Amines	148
ADRA1B	ADA1B	ADRA1B	a1B		Rhodopsin	Amines	147
ADRA1D	ADA1D	ADRA1D	a1D		Rhodopsin	Amines	146
ADRA2A	ADA2A	ADRA2A	a2A		Rhodopsin	Amines	150
ADRA2B	ADA2B	ADRA2B	a2B		Rhodopsin	Amines	151
ADRA2C	ADA2C	ADRA2C	a2C		Rhodopsin	Amines	152
ADRB1	ADRB1	ADRB1	b1		Rhodopsin	Amines	153
ADRB2	ADRB2	ADRB2	b2		Rhodopsin	Amines	154
ADRB3	ADRB3	ADRB3	b3		Rhodopsin	Amines	155
CHRM1	ACM1	CHRM1	M1		Rhodopsin	Amines	1128
CHRM2	ACM2	CHRM2	M2		Rhodopsin	Amines	1129
CHRM3	ACM3	CHRM3	M4		Rhodopsin	Amines	1131
CHRM4	ACM4	CHRM4	M3		Rhodopsin	Amines	1132
CHRM5	ACM5	CHRM5	M5		Rhodopsin	Amines	1133
DRD1	DRD1	DRD1	D1		Rhodopsin	Amines	1812
DRD2	DRD2	DRD2	D2		Rhodopsin	Amines	1813
DRD3	DRD3	DRD3	D3		Rhodopsin	Amines	1814
DRD4	DRD4	DRD4	D4		Rhodopsin	Amines	1815
DRD5	DRD5	DRD5	D5		Rhodopsin	Amines	1816
GPR57	TAAR3	GPR57			Rhodopsin	Amines	9288
GPR58	TAAR2	GPR58			Rhodopsin	Amines	9287
GPR61	GPR61	GPR61			Rhodopsin	Amines	83873
GPR62	GPR62	GPR62			Rhodopsin	Amines	118442
HRH1	HRH1	HRH1	H1		Rhodopsin	Amines	3269
HRH2	HRH2	HRH2	H2		Rhodopsin	Amines	3274
HRH3	HRH3	HRH3	H3		Rhodopsin	Amines	11255
HRH4	HRH4	HRH4	H4		Rhodopsin	Amines	59340
HTR1A	5HT1A	HTR1A	5-HT1A		Rhodopsin	Amines	3350
HTR1B	5HT1B	HTR1B	5-HT1B		Rhodopsin	Amines	3351

FamDBtool	UniProt	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
HTR1D	5HT1D	HTR1D	5-HT1D		Rhodopsin	Amines	3352
HTR1E	5HT1E	HTR1E	5-HT1E		Rhodopsin	Amines	3354
HTR1F	5H1TF	HTR1F	5-HT1F		Rhodopsin	Amines	3355
HTR2A	5HT2A	HTR2A	5-HT2A		Rhodopsin	Amines	3356
HTR2B	5HT2B	HTR2B	5-HT2B		Rhodopsin	Amines	3357
HTR2C	5HT2C	HTR2C	5-HT2C		Rhodopsin	Amines	3358
HTR4	5H4R	HTR4	5-HT4		Rhodopsin	Amines	3360
HTR5A	5HT5A	HTR5A	5-HT5A		Rhodopsin	Amines	3361
HTR6	5H6R	HTR6	5-HT6		Rhodopsin	Amines	3362
HTR7	5H7R	HTR7	5-HT7		Rhodopsin	Amines	3363
PNR	O14804	PNR			Rhodopsin	Amines	9038
TRAR1	TAR01	TRAR1		TA1	Rhodopsin	Amines	134864
TRAR3	TAR03	TRAR3		TA3	Rhodopsin	Amines	134860
TRAR4	TAR04	TRAR4		TA4	Rhodopsin	Amines	319100
TRAR5	TAR05	TRAR5		TA5	Rhodopsin	Amines	83551
GHSR	GHSR	GHSR	ghrelin		Rhodopsin	Brain-gut peptides	2693
GPR145	MCHR2	GPR145	MCH2		Rhodopsin	Brain-gut peptides	84539
GPR24	MCHR1	GPR24	MCH1		Rhodopsin	Brain-gut peptides	2847
GPR39	GPR39	GPR39			Rhodopsin	Brain-gut peptides	2863
MLNR	MTLR	MLNR	motilin		Rhodopsin	Brain-gut peptides	2862
NMUR1	Q9HB89	NMUR1	NMU1		Rhodopsin	Brain-gut peptides	10316
NMUR2	Q9GZQ4	NMUR2	NMU2		Rhodopsin	Brain-gut peptides	56923
NTSR1	NTR1	NTSR1	NTS1		Rhodopsin	Brain-gut peptides	4923
NTSR2	NTR2	NTSR2	NTS2		Rhodopsin	Brain-gut peptides	23620
TRHR	TRFR	TRHR	TRH		Rhodopsin	Brain-gut peptides	7201
AGTR1	AG2S	AGTR1	AT1		Rhodopsin	Chemoattractants	185
AGTR1	AG2R	AGTR1	AT1		Rhodopsin	Chemoattractants	185
AGTR2	AG22	AGTR2	AT2		Rhodopsin	Chemoattractants	186
AGTRL1	APJ	AGTRL1			Rhodopsin	Chemoattractants	187
BDKRB1	BKRB1	BDKRB1	B1		Rhodopsin	Chemoattractants	623
BDKRB2	BKRB2	BDKRB2	B2		Rhodopsin	Chemoattractants	624
C3AR1	C3AR	C3AR1			Rhodopsin	Chemoattractants	719
C5R1	C5AR	C5R1			Rhodopsin	Chemoattractants	728
CMKLR1	CML1	CMKLR1			Rhodopsin	Chemoattractants	1240
FPR1	FPR1	FPR1			Rhodopsin	Chemoattractants	2357
FPRL1	FPRL1	FPRL1	ALX		Rhodopsin	Chemoattractants	2358
FPRL2	FPR2	FPRL2			Rhodopsin	Chemoattractants	2359
GPR1	GPR1	GPR1			Rhodopsin	Chemoattractants	2825
GPR15	GPR15	GPR15			Rhodopsin	Chemoattractants	2838
GPR25	GPR25	GPR25			Rhodopsin	Chemoattractants	2848
GPR44	GPR44	GPR44			Rhodopsin	Chemoattractants	11251
GPR77	C5ARL	GPR77			Rhodopsin	Chemoattractants	27202
GPR146		GPR146			Rhodopsin	Chemokines	115330
CCBP2	CCBP2				Rhodopsin	Chemokines	1238

FamDBtool	UniProt	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
CCRL1	CCRL1				Rhodopsin	Chemokines	51554
ADMR	ADMR	ADMR			Rhodopsin	Chemokines	11318
BLR1	CXCR5	BLR1	CXCR5		Rhodopsin	Chemokines	643
CCR1	CCR1	CCR1	CCR1		Rhodopsin	Chemokines	1230
GPR2	CCR10	CCR10	CCR10		Rhodopsin	Chemokines	2826
CCR2	CCR2	CCR2	CCR2		Rhodopsin	Chemokines	1231
CCR3	CCR3	CCR3	CCR3		Rhodopsin	Chemokines	1232
CCR4	CCR4	CCR4	CCR4		Rhodopsin	Chemokines	1233
CCR5	CCR5	CCR5	CCR5		Rhodopsin	Chemokines	1234
CCR6	CCR6	CCR6	CCR6		Rhodopsin	Chemokines	1235
CCR7	CCR7	CCR7	CCR7		Rhodopsin	Chemokines	1236
CCR8	CCR8	CCR8	CCR8		Rhodopsin	Chemokines	1237
CCR9	CCR9	CCR9	CCR9		Rhodopsin	Chemokines	10803
CCRL2	O75307	CCRL2			Rhodopsin	Chemokines	9034
CMKOR1	RDC1	CMKOR1			Rhodopsin	Chemokines	57007
CX3CR1	C3X1	CX3CR1	CX3CR1		Rhodopsin	Chemokines	1524
CXCR3	CXCR3	CXCR3	CXCR3		Rhodopsin	Chemokines	2833
CXCR4	CXCR4	CXCR4	CXCR4		Rhodopsin	Chemokines	7852
CXCR6	CXCR6	CXCR6			Rhodopsin	Chemokines	10663
IL8RA	CXCR1	IL8RA	CXCR1		Rhodopsin	Chemokines	3577
IL8RB	CXCR2	IL8RB	CXCR2		Rhodopsin	Chemokines	3579
XCR1	XCR1	XCR1	XCR1		Rhodopsin	Chemokines	2829
GPR88	GPR88	GPR88			Rhodopsin	Glutamate	54112
FSHR	FSHR	FSHR			Rhodopsin	Glycoproteins	2492
GPR48	LGR4	LGR4			Rhodopsin	Glycoproteins	55366
GPR49	LGR5	LGR5			Rhodopsin	Glycoproteins	8549
LGR6	LGR6	LGR6			Rhodopsin	Glycoproteins	59352
LGR7	LGR7	LGR7			Rhodopsin	Glycoproteins	59350
LGR8	LGR8	LGR8			Rhodopsin	Glycoproteins	122042
LHCGR	LSHR	LHCGR			Rhodopsin	Glycoproteins	3973
TSHR	TSHR	TSHR			Rhodopsin	Glycoproteins	7253
CNR1	CNR1	CNR1	CB1		Rhodopsin	Lipids	1268
CNR2	CNR2	CNR2	CB2		Rhodopsin	Lipids	1269
EDG1	EDG1	EDG1	S1P1		Rhodopsin	Lipids	1901
EDG2	EDG2	EDG2	LPA1		Rhodopsin	Lipids	1902
EDG3	EDG3	EDG3	S1P3		Rhodopsin	Lipids	1903
EDG4	EDG4	EDG4	LPA2		Rhodopsin	Lipids	9170
EDG5	EDG5	EDG5	S1P2		Rhodopsin	Lipids	9294
EDG6	EDG6	EDG6	S1P4		Rhodopsin	Lipids	8698
EDG7	EDG7	EDG7	LPA3		Rhodopsin	Lipids	23566
EDG8	EDG8	EDG8	S1P5		Rhodopsin	Lipids	53637
GPR119	GPR119	GPR119			Rhodopsin	Lipids	139760
GPR12	GPR12	GPR12			Rhodopsin	Lipids	2835
GPR3	GPR3	GPR3			Rhodopsin	Lipids	2827

FamDBtool	UniProt	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GPR6	GPR6	GPR6			Rhodopsin	Lipids	2830
MRGX4	MRGX4	MRGX4			Rhodopsin	MA+D215	117196
MAS1	MAS	MAS1			Rhodopsin	MAS	4142
MAS1L	MAS1L	MAS1L			Rhodopsin	MAS	116511
MRGPRD	MRGPD	MRGPRD			Rhodopsin	MAS	116512
MRGE	MRGPE	MRGPRE			Rhodopsin	MAS	116534
MRGPRF	MRGPF	MRGPRF			Rhodopsin	MAS	219928
MRGPRG		MRGPRG			Rhodopsin	MAS	386746
MRGX1	MRGX1	MRGX1			Rhodopsin	MAS	259249
MRGX2	MRGX2	MRGX2			Rhodopsin	MAS	117194
MRGX3	MRGX3	MRGX3			Rhodopsin	MAS	117195
MC1R	MSHR	MC1R	MC1		Rhodopsin	Melanocortine	4157
MC2R	ACTHR	MC2R	MC2		Rhodopsin	Melanocortine	4158
MC3R	MC3R	MC3R	MC3		Rhodopsin	Melanocortine	4159
MC4R	MC4R	MC4R	MC4		Rhodopsin	Melanocortine	4160
MC5R	MC5R	MC5R	MC5		Rhodopsin	Melanocortine	4161
GPR22	GPR22	GPR22			Rhodopsin	Melatonin	2845
GPR45	GPR45	GPR45			Rhodopsin	Melatonin	11250
GPR50	MT1RL	GPR50			Rhodopsin	Melatonin	9248
GPR63	GPR63	GPR63			Rhodopsin	Melatonin	81491
MTNR1A	MTR1A	MTNR1A		MT1	Rhodopsin	Melatonin	4543
MTNR1B	MTR1B	MTNR1B		MT2	Rhodopsin	Melatonin	4544
GPR109A	Q8TDS4	GPR109A			Rhodopsin	Miscellaneous	338442
GPR109B	G109B	GPR109B			Rhodopsin	Miscellaneous	8843
GPR31	GPR31	GPR31			Rhodopsin	Miscellaneous	2853
GPR32	GPR32	GPR32			Rhodopsin	Miscellaneous	2854
GPR81	GPR81	GPR81			Rhodopsin	Miscellaneous	27198
OXER1	Q8TDS5	OXER1	OXE		Rhodopsin	Miscellaneous	165140
UTS2R	UR2R	UTS2R	UT		Rhodopsin	Miscellaneous	2837
PGR10		GPR149			Rhodopsin	n.d.	344758
FY	DUFFY				Rhodopsin	n.d.	2532
GPBAR1	Q8TDU6	GPBAR1			Rhodopsin	n.d.	151306
GPR120	GP120	GPR120			Rhodopsin	n.d.	338557
GPR135	GP135	GPR135			Rhodopsin	n.d.	64582
PGR3	Q6DWJ6	GPR139			Rhodopsin	n.d.	124274
GPR141	GP141	GPR141			Rhodopsin	n.d.	353345
GPR142	GP142	GPR142			Rhodopsin	n.d.	350383
GPR150	GP150	GPR150			Rhodopsin	n.d.	285601
GPR151	GP151	GPR151			Rhodopsin	n.d.	134391
GPR152	GP152	GPR152			Rhodopsin	n.d.	390212
GPR153	Q6NV75	GPR153			Rhodopsin	n.d.	387509
GPR160	GP160	GPR160			Rhodopsin	n.d.	26996
GPR18	GPR18	GPR18			Rhodopsin	n.d.	2841
GPR20	GPR20	GPR20			Rhodopsin	n.d.	2843

FamDBtool	UniProt	HUGO	Official IUPHAR nomen- clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GPR21	GPR21	GPR21			Rhodopsin	n.d.	2844
GPR26	GPR26	GPR26			Rhodopsin	n.d.	2849
GPR30	CML2	GPR30			Rhodopsin	n.d.	2852
GPR35	GPR35	GPR35			Rhodopsin	n.d.	2859
GPR37	GPR37	GPR37			Rhodopsin	n.d.	2861
GPR37L1	ETBR2	GPR37L1			Rhodopsin	n.d.	9283
GPR52	GPR52	GPR52			Rhodopsin	n.d.	9293
GPR55	GPR55	GPR55			Rhodopsin	n.d.	9290
GPR75	O95800	GPR75			Rhodopsin	n.d.	10936
GPR78	GPR78	GPR78			Rhodopsin	n.d.	27201
GPR82	GPR82	GPR82			Rhodopsin	n.d.	27197
GPR84	Q9NQS5	GPR84			Rhodopsin	n.d.	53831
GRCA	Q16538	GRCAb			Rhodopsin	n.d.	27239
LTB4R	LT4R1	LTB4R	BLT1		Rhodopsin	n.d.	1241
LTB4R2	LT4R2	LTB4R2	BLT2		Rhodopsin	n.d.	56413
GPR7	GPR7	GPR7			Rhodopsin	Opiates	2831
GPR8	GPR8	GPR8			Rhodopsin	Opiates	2832
OPRD1	OPRD	OPRD1	d		Rhodopsin	Opiates	4985
OPRK1	OPRK	OPRK1	k		Rhodopsin	Opiates	4986
OPRL1	OPRX	OPRL1	NOP		Rhodopsin	Opiates	4987
OPRM1	OPRM	OPRM1	m		Rhodopsin	Opiates	4988
SALPR	R3R1	RLN3R1			Rhodopsin	Opiates	51289
GPR100	R3R2	RLN3R2			Rhodopsin	Opiates	339403
SSTR1	SSR1	SSTR1	sst1		Rhodopsin	Opiates	6751
SSTR2	SSR2	SSTR2	sst2		Rhodopsin	Opiates	6752
SSTR3	SSR3	SSTR3	sst3		Rhodopsin	Opiates	6753
SSTR4	SSR4	SSTR4	sst4		Rhodopsin	Opiates	6754
SSTR5	SSR5	SSTR5	sst5		Rhodopsin	Opiates	6755
OPN1LW	OPSR				Rhodopsin	Opsins	5956
OPN1MW	OPSG				Rhodopsin	Opsins	2652
OPN1SW	OPSB				Rhodopsin	Opsins	611
OPN4	OPN4				Rhodopsin	Opsins	94233
RGR	RGR				Rhodopsin	Opsins	5995
RHO	OPSD				Rhodopsin	Opsins	6010
RRH	OPSX				Rhodopsin	Opsins	10692
OPN3	OPN3	OPN3			Rhodopsin	Opsins	23596
OPN5	Q6U736	OPN5			Rhodopsin	Opsins	221391
NPY6R					Rhodopsin	Peptides	4888
BRS3	BRS3	BRS3	BB3		Rhodopsin	Peptides	680
CCKAR	CCKAR	CCKAR	CCK1		Rhodopsin	Peptides	886
CCKBR	GASR	CCKBR	CCK2		Rhodopsin	Peptides	887
EDNRA	EDNRA	EDNRA	ETA		Rhodopsin	Peptides	1909
EDNRB	EDNRB	EDNRB	ETB		Rhodopsin	Peptides	1910
GALR1	GALR1	GALR1	GAL1		Rhodopsin	Peptides	2587

FamDBtool	UniProt	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GALR2	GALR2	GALR2	GAL2		Rhodopsin	Peptides	8811
GALR3	GALR3	GALR3	GAL3		Rhodopsin	Peptides	8484
GPR10	GPR10	GPR10	PRP		Rhodopsin	Peptides	2834
GPR103	QRFPR	GPR103			Rhodopsin	Peptides	84109
GPR147	NPFF1	GPR147			Rhodopsin	Peptides	64106
GPR19	GPR19	GPR19			Rhodopsin	Peptides	2842
GPR54	KISSR	GPR54			Rhodopsin	Peptides	84634
GPR74	NPFF2	GPR74			Rhodopsin	Peptides	10886
GPR83	GPR83	GPR83			Rhodopsin	Peptides	10888
GRPR	GRPR	GRPR	BB2		Rhodopsin	Peptides	2925
HCRTR1	OX1R	HCRTR1	OX1		Rhodopsin	Peptides	3061
HCRTR2	OX2R	HCRTR2	OX2		Rhodopsin	Peptides	3062
NMBR	NMBR	NMBR	BB1		Rhodopsin	Peptides	4829
NPY1R	NPY1R	NPY1R	Y1		Rhodopsin	Peptides	4886
NPY2R	NPY2R	NPY2R	Y2		Rhodopsin	Peptides	4887
NPY5R	NPY5R	NPY5R	Y5		Rhodopsin	Peptides	4889
PPYR1	NPY4R	PPYR1	Y4		Rhodopsin	Peptides	5540
TACR1	NK1R	TACR1	NK1		Rhodopsin	Peptides	6869
TACR2	NK2R	TACR2	NK2		Rhodopsin	Peptides	6865
TACR3	NK3R	TACR3	NK3		Rhodopsin	Peptides	6870
PTGDR	PD2R	PTGDR	DP		Rhodopsin	Prostanoid	5729
PTGER1	PE2R1	PTGER1	EP1		Rhodopsin	Prostanoid	5731
PTGER2	PE2R2	PTGER2	EP2		Rhodopsin	Prostanoid	5732
PTGER3	PE2R3	PTGER3	EP3		Rhodopsin	Prostanoid	5733
PTGER4	PE2R4	PTGER4	EP4		Rhodopsin	Prostanoid	5734
PTGFR	PF2R	PTGFR	FP		Rhodopsin	Prostanoid	5737
PTGIR	PI2R	PTGIR	IP1		Rhodopsin	Prostanoid	5739
TBXA2R	TA2R	TBXA2R	TP		Rhodopsin	Prostanoid	6915
5-HT5B-like					Rhodopsin	pseudogene	343958
HTR7-like					Rhodopsin	pseudogene	93164
GPR33		GPR33			Rhodopsin	pseudogene	2856
CYSLTR1	CLTR1	CYSLTR1	CysLT1		Rhodopsin	Purines	10800
CYSLTR2	CLTR2	CYSLTR2	CysLT2		Rhodopsin	Purines	57105
EBI2	EBI2	EBI2			Rhodopsin	Purines	1880
F2R	PAR1	F2R	PAR1		Rhodopsin	Purines	2149
F2RL1	PAR2	F2RL1	PAR2		Rhodopsin	Purines	2150
F2RL2	PAR3	F2RL2	PAR3		Rhodopsin	Purines	2151
F2RL3	PAR4	F2RL3	PAR4		Rhodopsin	Purines	9002
GPR132	G2A	GPR132			Rhodopsin	Purines	29933
GPR17	GPR17	GPR17			Rhodopsin	Purines	2840
GPR171	GP171	GPR171			Rhodopsin	Purines	29909
FKSG79	GP174	GPR174			Rhodopsin	Purines	84636
GPR23	P2RY9	GPR23			Rhodopsin	Purines	2846
GPR34	GPR34	GPR34			Rhodopsin	Purines	2857

FamDBtool	UniProt	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GPR4	GPR4	GPR4			Rhodopsin	Purines	2828
GPR40	GPR40	GPR40			Rhodopsin	Purines	2864
GPR41	GPR41	GPR41			Rhodopsin	Purines	2865
GPR42	GPR42	GPR42			Rhodopsin	Purines	2866
GPR43	GPR43	GPR43			Rhodopsin	Purines	2867
GPR65	PSYR	GPR65			Rhodopsin	Purines	8477
GPR68	SPR1	GPR68			Rhodopsin	Purines	8111
P2Y3L		GPR79			Rhodopsin	Purines	27200
GPR87	GPR87	GPR87			Rhodopsin	Purines	53836
GPR92	GPR92	GPR92			Rhodopsin	Purines	57121
OXGR1	GPR80	OXGR1			Rhodopsin	Purines	27199
P2RY1	P2RY1	P2RY1	P2Y1		Rhodopsin	Purines	5028
P2RY10	P2Y10	P2RY10			Rhodopsin	Purines	27334
P2RY11	P2Y11	P2RY11	P2Y11		Rhodopsin	Purines	5032
P2RY12	P2Y12	P2RY12		P2Y12	Rhodopsin	Purines	64805
GPR86	GPR86	P2RY13		P2Y13	Rhodopsin	Purines	53829
GPR105	P2Y14	P2RY14			Rhodopsin	Purines	9934
P2RY2	P2RY2	P2RY2	P2Y2		Rhodopsin	Purines	5029
P2RY4	P2RY4	P2RY4	P2Y4		Rhodopsin	Purines	5030
P2RY5	P2RY5	P2RY5			Rhodopsin	Purines	10161
P2RY6	P2RY6	P2RY6	P2Y6		Rhodopsin	Purines	5031
P2RY8		P2RY8			Rhodopsin	Purines	286530
PTAFR	PTAFR	PTAFR	PAF		Rhodopsin	Purines	5724
SUCNR1	GPR91	SUCNR1			Rhodopsin	Purines	56670
GPR	Q14439	GPR			Rhodopsin	SREBs	11245
GPR101	GP101	GPR101			Rhodopsin	SREBs	83550
GPR161	GP161	GPR161			Rhodopsin	SREBs	23432
SREB3	GP173	GPR173			Rhodopsin	SREBs	54328
GPR27	GPR27	GPR27			Rhodopsin	SREBs	2850
GPR85	GPR85	GPR85			Rhodopsin	SREBs	54329
AVPR1A	V1AR	AVPR1A	V1A		Rhodopsin	Vasopeptides	552
AVPR1B	V1BR	AVPR1B	V1B		Rhodopsin	Vasopeptides	553
AVPR2	V2R	AVPR2	V2		Rhodopsin	Vasopeptides	554
GPR154	Q6W5P4	GPR154			Rhodopsin	Vasopeptides	387129
GPR73	PKR1	GPR73	PK1		Rhodopsin	Vasopeptides	10887
GPR73L1	PKR2	GPR73L1	PK2		Rhodopsin	Vasopeptides	128674
OXTR	OXYR	OXTR	OT		Rhodopsin	Vasopeptides	5021

FamDBtool	Rognan-DB	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GPR157		GPR157			Secretin	Secretin	80045
ADCYAP1R1	PACR	ADCYAP1R1	PAC1		Secretin	Secretin	117
CALCR	CALCR	CALCR	CT		Secretin	Secretin	799
CALCRL	CALRL	CALCRL	CGRP1		Secretin	Secretin	10203
CRHR1	CRFR1	CRHR1	CRF1		Secretin	Secretin	1394
CRHR2	CRFR2	CRHR2	CRF2		Secretin	Secretin	1395
GCGR	GLR	GCGR	glucagon		Secretin	Secretin	2642
GHRHR	GHRHR	GHRHR	GHRH		Secretin	Secretin	2692
GIPR	GIPR	GIPR	GIP		Secretin	Secretin	2696
GLP1R	GLP1R	GLP1R	GLP-1		Secretin	Secretin	2740
GLP2R	GLP2R	GLP2R	GLP-2		Secretin	Secretin	9340
PTHR1	PTHR1	PTHR1		PTH1	Secretin	Secretin	5745
PTHR2	PTHR2	PTHR2		PTH2	Secretin	Secretin	5746
SCTR	SCTR	SCTR	secretin		Secretin	Secretin	6344
VIPR1	VIPR1	VIPR1	VPAC1		Secretin	Secretin	7433
VIPR2	VIPR2	VIPR2	VPAC2		Secretin	Secretin	7434
GPR133	GP133				Adhesion	Adhesion	283383
BAI1	BAI1	BAI1			Adhesion	Adhesion	575
BAI2	BAI2	BAI2			Adhesion	Adhesion	576
BAI3	BAI3	BAI3			Adhesion	Adhesion	577
CD97	CD97	CD97			Adhesion	Adhesion	976
CELSR1	CELR1	CELSR1			Adhesion	Adhesion	9620
CELSR2	CELR2	CELSR2			Adhesion	Adhesion	1952
CELSR3	CELR3	CELSR3			Adhesion	Adhesion	1951
ELTD1	ELTD1	ELTD1			Adhesion	Adhesion	64123
EMR1	EMR1	EMR1			Adhesion	Adhesion	2015
EMR2	EMR2	EMR2			Adhesion	Adhesion	30817
EMR3	EMR3	EMR3			Adhesion	Adhesion	84658
GPR127	EMR4	EMR4			Adhesion	Adhesion	326342
GPR110	GP110	GPR110			Adhesion	Adhesion	266977
GPR111	GP111	GPR111			Adhesion	Adhesion	222611
GPR112	GP112	GPR112			Adhesion	Adhesion	139378
GPR113	GP113	GPR113			Adhesion	Adhesion	165082
GPR114	GP114	GPR114			Adhesion	Adhesion	221188
GPR115	GP115	GPR115			Adhesion	Adhesion	221393
GPR116	GP116	GPR116			Adhesion	Adhesion	221395
GPR123	GP123	GPR123			Adhesion	Adhesion	84435
GPR124	GP124	GPR124			Adhesion	Adhesion	25960
GPR125	GP125	GPR125			Adhesion	Adhesion	166647
GPR126	GP126	GPR126			Adhesion	Adhesion	57211
GPR128	GP128	GPR128			Adhesion	Adhesion	84873
GPR144	GP144	GPR144			Adhesion	Adhesion	347088
GPR56	GPR56	GPR56			Adhesion	Adhesion	9289
GPR64	GPR64	GPR64			Adhesion	Adhesion	10149

FamDBtool	Rognan-DB	HUGO	Official IUPHAR nomen-clature	Provisional name IUPHAR	Class	Cluster	Locus ID
GPR97	GPR97	GPR97			Adhesion	Adhesion	222487
LPHN1	LPHN1	LPHN1			Adhesion	Adhesion	22859
LPHN2	LPHN2	LPHN2			Adhesion	Adhesion	23266
LPHN3	LPHN3	LPHN3			Adhesion	Adhesion	23284
MASS1	Q8WXG9	MASS1			Adhesion	Adhesion	84059
GPR143	GP143				Adhesion	n.d.	4935
GPR158L1		GPR158L1			Glutamate	Glutamate	440435
LOC344760	Q8TDU1				Glutamate	Glutamate	344760
TAS1R1	TS1R1				Glutamate	Glutamate	80835
TAS1R2	TS1R2				Glutamate	Glutamate	80834
TAS1R3	TS1R3				Glutamate	Glutamate	83756
CASR	CASR	CASR		CaS	Glutamate	Glutamate	846
GABBR1	GABR1	GABBR1	GABAB1		Glutamate	Glutamate	2550
RAI3	RAI3	GPCR5A		RAIG1	Glutamate	Glutamate	9052
GPR156	Q8NFN8	GPR156		GABABL	Glutamate	Glutamate	165829
GPR158	Q6QR81	GPR158			Glutamate	Glutamate	57512
GPR51	GABR2	GPR51	GABAB2		Glutamate	Glutamate	9568
GPRC5B	GPC5B	GPRC5B		RAIG2	Glutamate	Glutamate	51704
GPRC5C	GPC5C	GPRC5C		RAIG3	Glutamate	Glutamate	55890
GPRC5D	GPC5D	GPRC5D			Glutamate	Glutamate	55507
GPRC6A	Q8NHZ9	GPRC6A			Glutamate	Glutamate	222545
GRM1	MGR1	GRM1	mGlu1		Glutamate	Glutamate	2911
GRM2	MGR2	GRM2	mGlu2		Glutamate	Glutamate	2912
GRM3	MGR3	GRM3	mGlu3		Glutamate	Glutamate	2913
GRM4	MGR4	GRM4	mGlu4		Glutamate	Glutamate	2914
GRM5	MGR5	GRM5	mGlu5		Glutamate	Glutamate	2915
GRM6	MGR6	GRM6	mGlu6		Glutamate	Glutamate	2916
GRM7	MGR7	GRM7	mGlu7		Glutamate	Glutamate	2917
GRM8	MGR8	GRM8	mGlu8		Glutamate	Glutamate	2918
FZD1	FZD1	FZD1		FZD1	Frizzled	Frizzled	8321
FZD10	FZ10	FZD10		FZD10	Frizzled	Frizzled	11211
FZD2	FZD2	FZD2		FZD2	Frizzled	Frizzled	2535
FZD3	FZD3	FZD3		FZD3	Frizzled	Frizzled	7976
FZD4	FZD4	FZD4		FZD4	Frizzled	Frizzled	8322
FZD5	FZD5	FZD5		FZD5	Frizzled	Frizzled	7855
FZD6	FZD6	FZD6		FZD6	Frizzled	Frizzled	8323
FZD7	FZD7	FZD7		FZD7	Frizzled	Frizzled	8324
FZD8	FZD8	FZD8		FZD8	Frizzled	Frizzled	8325
FZD9	FZD9	FZD9		FZD9	Frizzled	Frizzled	8326
SMO	SMO	SMO		SMO	Frizzled	Frizzled	6608

3
A Novel Drug Screening Assay for G Protein-coupled Receptors

Brian F. O'Dowd, Xiaodong Ji, Mohammad Alijaniaram, Tuan Nguyen,
and Susan R. George

3.1
Introduction

3.1.1
History

GPCRs are proven targets for drug discovery [1], and are by far the largest subgroup of cell surface receptors that transduce signals for molecules such as monoamines, peptides, glycoprotein hormones and the large class of odorant molecules [2]. The range of therapeutic areas targeted by GPCRs is diverse and the commercial potential of these receptors is not yet fully realized. For instance, among the ~367 GPCRs (not including odorant receptors), there are 195 for which the endogenous ligands are known, but only 30 of these have been targeted by drugs in use presently [3]. GPCR drugs are estimated to represent 40–50 % of the current prescription market [4].

The remaining ~170 GPCRs are orphan receptors, for which the endogenous ligands are as yet unidentified [2, 5]. The recently decoded orphan GPCRs illustrate the enormous untapped potential represented by these receptors for drug discovery [5, 6]. Although various methods have hastened the discovery of the genes encoding GPCRs, the identification of their endogenous ligands and novel compounds targeting these receptors has lagged far behind and there is a critical need for highly effective novel approaches [7]. As the signal transduction pathways involved in the activation of an orphan GPCR cannot be predicted, an assay system which is independent of the G protein employed by the receptor is highly desirable. Although several methods are currently available to discover the endogenous ligands and screen for drugs targeting GPCRs [8–10], no single method has surpassed the capabilities of the others. Many of the present screening assays in use have varying requirements for specific G protein coupling to generate a signal, or prior identification of an agonist in order to screen for antagonists. Among the methods that are currently in use, is one that uses the major principle of

β-arrestin translocation to detect agonist activation of the GPCR, and another that uses inhibition of constitutive receptor activity to detect antagonist action.

3.1.2
Nuclear Translocation of Endogenous GPCRs

In this chapter, we describe a novel method, which harnesses a well-characterized physiological cellular mechanism for transport of proteins to the nucleus. This transport pathway was adapted for the identification of chemicals blocking both known GPCRs and orphan GPCRs. The translocation of cytosolic proteins to the nucleus occurs by well-characterized mechanisms, often involving nuclear transport proteins that recognize distinct motifs called nuclear localization sequences (NLSs) [11], and mediate the translocation of "cargo" proteins to the nucleus. Few of the GPCRs contain NLSs [12–14]. GPCRs with a recognized nuclear localization are limited, with little evidence of known nuclear function as yet. One example, the angiotensin AT1 receptor contains an endogenous NLS, which reportedly serves to direct the receptor into the nucleus in response to an agonist [12]. Nuclear import of proteins requires the docking of the "cargo" protein to an importin α/β heterodimer. Importin β recognizes cargo proteins by binding to NLS sequences which usually consist of short stretches of basic amino acids, arginine and lysine [11].

We reported an agonist-independent nuclear localization of the angiotensin AT1, bradykinin B2 and apelin receptors [15]. The cytoplasmic tail of the AT1 receptor contains an NLS sequence KKFKR [12] downstream of the transmembrane domain 7 within helix 8. A similar NLS-like sequence is also present in helix 8 of the bradykinin B2 receptor. Thus the AT2 and the B2 receptor were observed to have similar NLS sequence situated within helix 8, located in the receptor carboxyl tails. The apelin receptor also demonstrated agonist-independent nuclear localization [15]. We reported that a sequence RKRRR is a functional NLS sequence located in the third intracellular loop of the apelin receptor. In total our search of the rhodopsin family GPCR sequences revealed 17 GPCRs with an NLS sequence in helix 8, including adenosine, growth hormone secretagogue, motilin, purinergic and orphan receptors in addition to the angiotensin, bradykinin and endothelin receptor [15].

3.1.3
The MOCA Method

We have now shown that genetically incorporating the NLS from the AT1 receptor into other GPCRs, introduced into a conformationally sensitive site, mediated a time-dependent and ligand-independent translocation of the GPCR from the cell surface to the nucleus. Since the optimal placement of the NLS provided conformational sensitivity, this engineered translocation pathway has been exploited to identify interacting compounds, which were able to modulate the ligand-independent transfer of the GPCR away from the cell membrane.

3.2
The MOCA Strategy Demonstrated with the D1 Dopamine Receptor

3.2.1
Development of the Assay

The dopamine D1 receptor was selected as the prototype membrane protein with which to investigate the MOCA experimental strategy. The optimal placement of the NLS was investigated by its introduction into various positions within the three intracellular loops and cytoplasmic tail of the D1 receptor. Multiple sites of incorporation were tested and from this we determined that the preferred placement of the NLS was in helix 8 of the D1 dopamine receptor, where conformation-dependent effects of ligand occupancy were demonstrated. One of the major advantages of the NLS-incorporation method is the ability to perform these studies in cells transiently expressing the receptor to be tested, allowing for rapid, high throughput screening without the delays associated with the generation of cell lines stably expressing the proteins.

The D1 dopamine receptor with an NLS inserted in the first intracellular cytoplasmic loop (D1-IC1-NLS-GFP) was expressed and detected in the nucleus of 85 % of cells. Treatment with either of the antagonists (+)butaclamol (1 µM) or SCH-23390 (1 µM) resulted in retention of the receptor at the cell surface in 82 and 77 % of the cells respectively, compared to receptor expression in the nucleus in 76 % of cells undergoing vehicle treatment.

The D1 dopamine receptor with an NLS inserted in the second intracellular cytoplasmic loop (D1-IC2-NLS-GFP) was expressed and detected in the nucleus of 51 % of cells. However over 40 % of cells had detectable receptors on the cell surface indicating that incorporation of the NLS at this position was not as efficient in translocating the receptor from the cell surface.

The D1 dopamine receptor with an NLS inserted in the third intracellular cytoplasmic loop (D1-IC3-NLS-GFP) was detected in the nucleus of 85 % of cells. However, treatment with various concentrations of SCH-23390 revealed ~84 % of cells with receptors in the nucleus, indicating that there was little response to the drug treatment. We investigated the role of the incorporated NLS further by analyzing the engineered NLS after it was mutated or shifted, in order to prove that this sequence mediated the nuclear transfer. We mutated the NLS sequence KKFKR to KKAKR in the D1-NLS receptor, and found a marked reduction in the ability of the receptor to translocate to the nucleus.

When expressed in cells, the D1 receptor tagged with GFP (D1-GFP) was localized exclusively on the cell surface at 48 h post-transfection as determined by confocal microscopy. The majority, > 90 %, of the cells demonstrated cell surface expression of the receptor (Fig. 3.1 a). Cells transfected with the D1 receptor containing an NLS in helix 8 at the base of transmembrane domain 7 (D1-NLS-GFP), revealed a time-dependent appearance and localization of the receptor on the cell surface (9 to 10 h post-transfection) and then translocation to the nucleus (24–48 h post-transfection) was observed in the majority of cells, with no
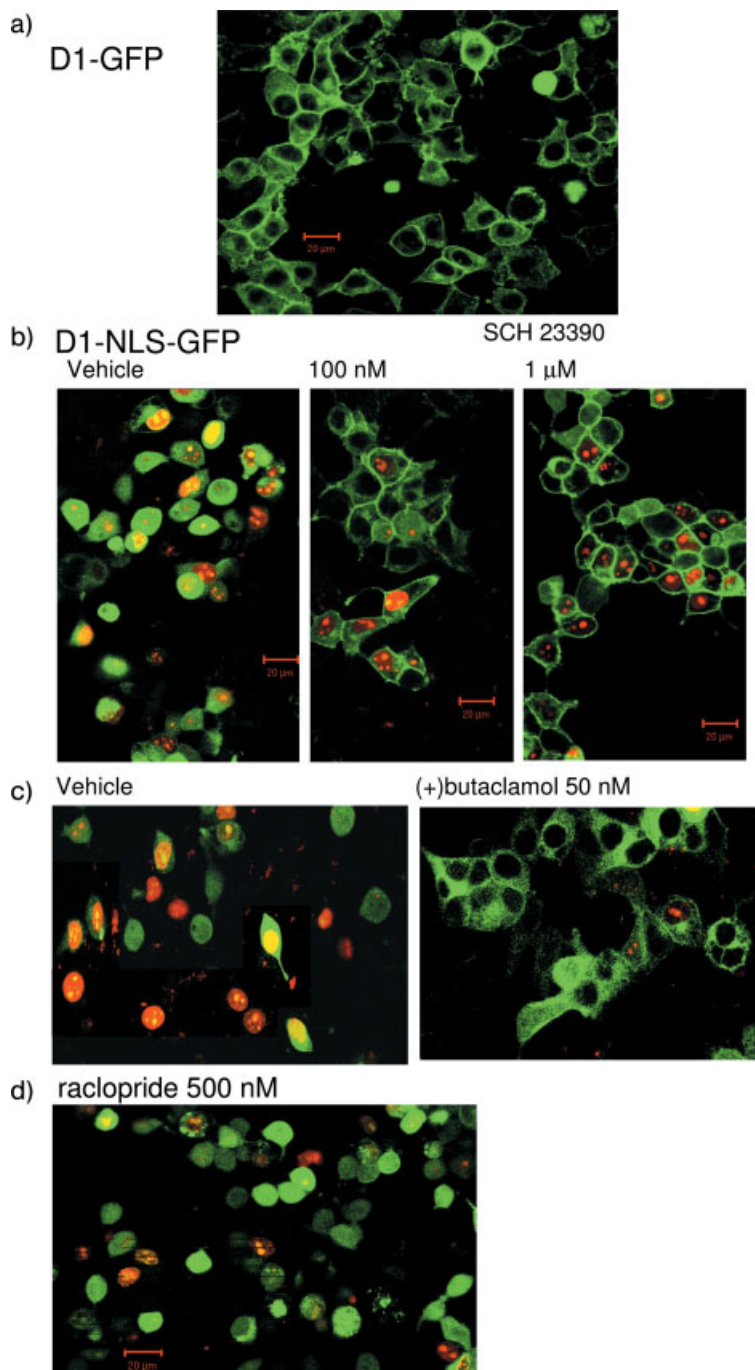
Fig. 3.1 (legend see page 55)

cell surface receptors visible (Fig. 3.1 b, left panel). Thus, incorporation of an NLS into the D1 receptor sequence in helix 8 resulted in a very efficient removal of the receptor from the cell surface under basal conditions with trafficking to the nucleus.

3.2.2
Concentration-dependent Antagonist Blockade of Nuclear Transport

Cells expressing D1-NLS-GFP were treated with the antagonists SCH-23390 or (+)butaclamol in varying concentrations. Control cells received only vehicle treatment. With SCH-23390 (10 nM to 10 µM) or (+)butaclamol (10 nM to 10 µM) treatment for 48 h, there was a dose-dependent retention of the D1-NLS-GFP receptor on the cell surface with marked diminution of the translocation to the nucleus (see examples of drug-treated cells in Fig. 3.1 b and c). Stereoselectivity for the drug effect was demonstrated by the lack of effect of (–)butaclamol (1 µM), in contrast to the robust effects of equivalent concentrations of (+)butaclamol to inhibit nuclear translocation of D1-NLS-GFP. Thus, treatment with D1-selective antagonists efficiently prevented, in a dose-responsive manner with nanomolar effectiveness, the nuclear translocation of the NLS-containing receptor.

The specificity of the antagonist effect on receptor translocation was tested by treating cells expressing D1-NLS-GFP with the D2-selective antagonist raclopride (500 nM). Following this treatment, D1-NLS-GFP was present in 87% of cell nuclei and in 0% at the cell surface, indicating that raclopride was unable to retain D1-NLS-GFP on the cell surface.

3.2.3
Measurement of Receptor Cell Surface Expression: Antagonist Binding of Receptors at Cell Surface

Radioligand binding was used to provide a direct quantification of the D1-NLS receptor density in the plasma membrane, isolated by sucrose density gradient fractionation. Cells expressing D1-NLS were treated with a range of antagonist (+)butaclamol concentrations or vehicle. The receptor density on the cell surface was quantified by specific binding of [^{3}H]-SCH23390.

Fig. 3.1 Effect of incorporation of an NLS into helix 8 of the D1 dopamine receptor and treatment with dopamine antagonists. The receptor–GFP fusion protein was expressed in HEK cells and visualized by confocal microscopy. Live cells expressing GFP [17] fusion proteins were visualized with an LSM510 Zeiss confocal laser microscope. Nuclei were identified by co-expression with DsRed2-nuc. D1-GFP was localized largely on the cell surface and was absent from the nucleus (a). D1-NLS-GFP-expressing cells were treated with vehicle, the D1-selective antagonist SCH 23390 (b), the non-selective antagonist (+)butaclamol (c) and the D2-selective antagonist raclopride (d). Incorporation of the NLS into the receptor caused efficient removal of D1-NLS-GFP from the cell surface with localization in the nucleus. Treatment with the D1 antagonists prevented the translocation, whereas the D2 antagonist had no effect.

(+)Butaclamol Concentration

Fig. 3.2 Expression of the dopamine D1 receptor with an inserted NLS, treatment with antagonist and quantification of plasma membrane receptor density by radioligand binding. Cells expressing D1-NLS were treated with varying doses of (+)butaclamol or vehicle. Plasma membranes were harvested by sucrose density gradient and receptor density measured. D1-NLS left untreated was present on the cell surface in very low density and antagonist treatment resulted in a dose-dependent retention of receptor in the plasma membrane. Radioligand binding: cells expressing the membrane protein were treated with an antagonist for 48 h. Plasma membranes were isolated by discontinuous sucrose density gradient centrifugation and receptor density determined by radioligand binding of [^{3}H]-SCH 23390 with specific binding defined by (+)butaclamol 1 μM, as described [18]. The membranes were incubated with radioligand and varying drug concentrations for 2 h at room temperature and collected using a Brandell cell harvester for scintillation counting. The binding analyses were performed in duplicate and the experiment was repeated three times.

Untreated control cells expressing D1-NLS revealed plasma membrane receptor density of 1.42 ± pmol/mg protein; whereas in cells treated with (+)butaclamol (500 nM), the plasma membrane D1 receptor density was 13.15 ± pmol/mg protein. The effect of a range of concentrations (10 nM to 10 μM) of (+)butaclamol was tested and revealed a significant dose–response effect (Fig. 3.2).

3.3
Development of Quantitative Methodology Suitable for High Throughput Analysis

Quantitative analysis suitable for high throughput analysis was achieved by the use of fluorescence cell surface labeling with the entire method performed in multi-well plates with the signal detected by a plate-reader fluorimeter. Cells expressing HA-tagged D1-NLS receptors were treated with a (+)butaclamol concentrations (10 nM to 10 μM). There was a dose-dependent effect of SCH 23390 (Fig. 3.3 a) and (+)butaclamol (Fig. 3.3 b) to retain the receptor on the cell surface, indicating that these antagonists reduced receptor trafficking from the cell surface

a)

Cell Surface Fluorescence

EC50= 24.5 nM

Relative Fluorescence Units

Log SCH 23390 Concentration (nM)

b)

EC50= 29.3 nM

Relative Fluorescence Units

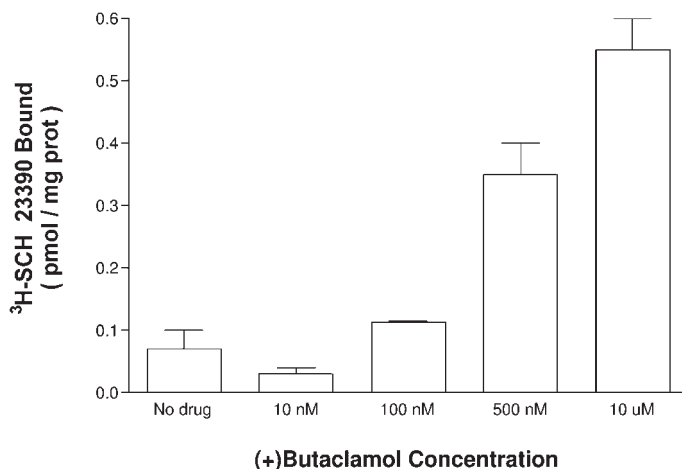Log (+)butaclamol Concentration (nM)
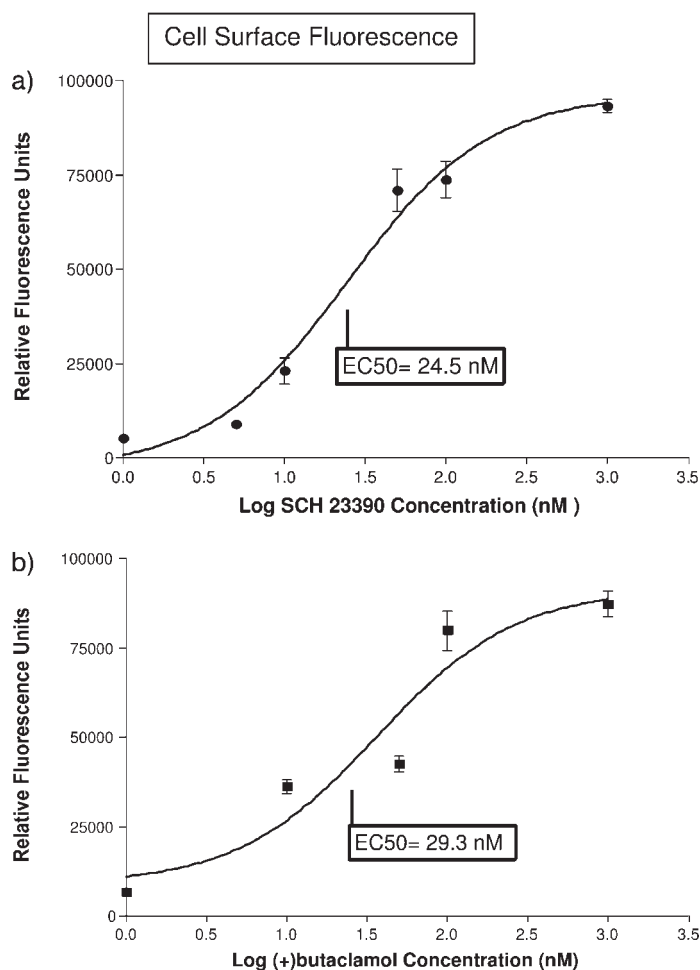
Fig. 3.3 Expression of the dopamine D1 receptor with an inserted NLS, treatment with the antagonist and quantification of cell surface receptor density by fluorimetric analysis. Cells expressing HA-tagged D1-NLS receptor were treated with antagonist and cell surface receptors quantified by the fluorescent signal of a FITC-conjugated second antibody directed at the HA antibody. Treatment with the antagonists SCH 23390 (a) and (+)butaclamol (b) revealed robust dose-dependent effects to retain HA-D1-NLS at the cell surface, indicating that the antagonists reduced translocation from the cell surface.

and fluorimetric analysis permitted detection and quantitation of the receptor retained at the cell surface and was adaptable to a multi-well format. In this method ~ 50,000 cells were added to each well (in a 96-well plate) and transfected with 0.5 µg cDNA. Medium containing the drug at varying concentrations was added to the wells. After 48 h, cells were fixed with paraformaldehyde and incubated with the primary antibody (rat anti-HA antibody) and secondary antibody conju-

gated to FITC (goat anti-rat antibody). Cell surface fluorescence was detected using a Cytofluor 4000 (PerSeptive Biosystems, Framingham, MA, USA).

3.3.1
Nuclear Translocation of Orphan GPCRs

An NLS was incorporated into helix 8 of GPR54, a recently deorphanized receptor that was first identified in our laboratory [16]. When expressed in cells, this modified receptor was efficiently translocated to the nucleus, in contrast to the wild-type GPR54 receptor which had a predominantly cell surface expression, demonstrating that this method can be used to screen for ligands targeting orphan receptors.

3.4
Discussion of the MOCA Method

The novel assay described and validated utilizing the D1 dopamine receptor has unique and original features, namely an agonist-independent translocation of the GPCR to the nucleus by the strategic introduction of an NLS into the protein sequence, with sequestration of the GPCR away from the cell surface and into the nucleus, likely by interaction with an importin protein. Blockade of this pathway was used as the basis for detection of the antagonists. The validation and reliability of the MOCA method for the GPCRs tested are features which predict that the use of this assay will contribute significantly to the process of drug discovery for GPCRs, including orphan GPCRs, for many of which there are no convenient and reliable drug screening methods. In our laboratory the MOCA strategy has demonstrated its versatility with respect to the multiple receptor detection systems employed and provides a convenient and simple to use tool for screening candidate compounds for their ability to interact with GPCRs.

For GPCRs, the assay provides several significant advances over the currently available assay methodologies. The method is applicable to drug discovery for GPCRs coupling to any type of G protein or interacting with different types of endogenous ligands and does not require any prior knowledge of the natural effector system or second messenger linkage of the GPCR, unlike the disadvantages associated with many commercially utilized GPCR drug discovery assays. The method is particularly suitable for orphan GPCRs.

The actions of the GPCR antagonists demonstrated sensitive dose–response effects at nanomolar concentrations and maintained specificity to subtype-selective compounds. Using the D1 dopamine receptor, we have demonstrated that the method is suitable for detecting antagonists ((+)butaclamol, and SCH23390), displayed stereoselectivity (lack of effect of (–)butaclamol) and a high degree of specificity, with a lack of effect of the D2 receptor antagonist raclopride.

This method has capitalized on the physiological mechanism whereby short sequences of amino acids, specifying an NLS, when incorporated into cytoplasmic proteins enable the proteins to be relocated to the nucleus [11], likely by the impor-

tin translocation pathway. Most GPCRs do not have an endogenous NLS, and predictably do not traffic to the nucleus in their unmodified states. Therefore, the incorporation of an NLS into a GPCR and its basal translocation off the cell surface into the nucleus has formed the basis for the assay. To take advantage of the high throughput potential, a fluorescence plate reader methodology was also developed. These different capabilities and properties within a single assay provide a novel platform that will be a convenient, rapid and easy to use tool that will provide a unique alternative within the spectrum of drug screening assays currently available.

3.5
Conclusion

The highly novel general strategy we have employed involves the robust translocation of the GPCR with a specific genetic modification from the cell surface in a basal time-dependent and ligand-independent manner with no recycling of the protein back to the cell surface. Interaction of the GPCR with structurally compatible compounds prevents translocation away from the cell surface and is measured as protein retained on the cell surface.

There is general recognition from pharmaceutical companies that a novel and fresh drug discovery platform would have a significant place in the drug discovery arena. The target users for MOCA will include the large pharmaceutical companies and numerous middle-ranged drug companies, many of whom have active GPCR drug discovery programs. The competitive advantage stems from the uniqueness of the assay product.

There is an untapped and untested number of potential drug targets and interest from the pharmaceutical companies to utilize these targets for drug development. We have described an entirely novel method for the identification of chemical compounds that block GPCRs. This method will not only lead to an acceleration of the process by which new drug molecules are recognized but will also pave the way for the discovery of therapeutically active drugs with potential medical and quality-of-life implications for all. There is a recognized need to have improved, versatile assay systems, where not just endogenous ligands, but synthetic novel ligands can be tested and identified in a quick and efficient manner that is amenable to automation.

Acknowledgments

References

1 Pierce, K.L., Premont, R.T., Lefkowitz, R.J. Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* **2002**; *3*: 639–650.

2 Lee, D.K., George, S.R., O'Dowd, B.F. Novel G-protein-coupled receptor genes expressed in the brain: continued discovery of important therapeutic targets. *Expert Opin Ther Targets* **2002**; *6*: 185–202.

3 Vassilatis, D.K, Hohmann, J.G, Zeng, H, Li, F, Ranchalis, J.E, Mortrud, M.T, Brown, A, Rodriguez, S.S.,Weller, J.R.,Wright A.C., Bergmann, J.E., Gaitanaris, G.A. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci USA* **2003**; *100*: 4903–4908.

4 Wise, A., Gearing, K., Rees, S. Target validation of G-protein coupled receptors. *Drug Disc. Today* **2002**; *7*: 235–246.

5 Bailey,W.J.,Vanti, W. B., George, S.R., Blevins, R., Swaminathan, S., Bonini, J.A., Smith, K.E.,Weinshank, R.L., O'Dowd, B.F. Patent status of the therapeutically important G-protein-coupled receptors. *Expert Opin Ther Patents* **2001**; *11*: 1861–1887.

6 Zambrowicz, B.P., Sands, A.T. Knockouts model the 100 best-selling drugs – will they model the next 100? *Nat Rev Drug Disc* **2003**; *2*: 38–50.

7 Howard, A.D., McAllister, G., Feighner, S.D., Liu, Q., Nargund, R.P., Van der Ploeg, L.H., Patchett, A.A. Orphan G-protein-coupled receptors and natural ligand discovery. *Trends Pharmacol Sci* **2001**; *22*: 132–140.

8 Szekeres, P.G. Functional assays for identifying ligands at orphan G protein-coupled receptors. *Recep Chann* **2002**; *8*: 297–308.

9 Barak, L.S., Ferguson, S.S.G., Zhang, Z., Caron, M.G. A β-Arrestin/green fluorescent protein biosensor for detecting G protein-coupled receptor activation. *J Biol Chem* **1997**; *272*: 27497–27500.

10 Chalmers, D.T., Behan, D.P. The use of constitutively active GPCRs in drug discovery and functional genomics. *Nat Rev Drug Disc* **2002**; *1*: 599–608.

11 Jans, D.A., Xiao, C-Y., Lam, M.H.C. Nuclear targeting signal recognition: a key control point in nuclear transport? *BioEssays* **2000**; *22*: 532–544.

12 Lu, D.,Yang, H., Shaw, G., Raizada, M.K. Angiotensin II-induced nuclear targeting of the angiotensin type 1 (AT1) receptor in brain neurons. *Endocrinology* **1998**; *139*: 365–375.

13 Chen, R., Mukhin,Y.V., Garnovskaya, M.N.,Thielen,T.E., Iijima,Y., Huang, C., Raymond, J.R., Ullian, M.E., Paul, R.V. A functional angiotensin II receptor–GFP fusion protein: evidence for agonist-dependent nuclear translocation. *Am J Physiol Renal Physiol* **2000**; *279*: F440–F448.

14 Watson, P.H., Fraher, L.J., Natale, B.V., Kisiel, M., Hendy, G.N., Hodsman, A.B. Nuclear localization of the type 1 parathyroid hormone/parathyroid hormone-related peptide receptor in MC3T3-E1 cells: association with serum-induced cell proliferation. *Bone* **2000**; *26*: 221–225.

15 Lee, D.K., Lanca, A.J., Cheng, R., Nguyen,T., Ji, X.D., Gobeil, F Jr., Chemtob, S., George, S.R., O'Dowd, B.F. Agonist-independent nuclear localization of the apelin, angiotensin AT1, and bradykinin B2 receptors. *J Biol Chem* **2004**; *279*: 7901–7908.

16 Lee, D.K., Nguyen,T., O'Neill, G.P., Cheng, R., Liu,Y., Howard, A.D., Coulombe, N.,Tan, C.P.,Tang-Nguyen, A.T., George, S.R., O'Dowd, BF. Discovery of a receptor related to the galanin receptors. *FEBS Lett* **1999**; *446*: 103–107.

17 Prasher, D.C., Eckenrode,V.K.,Ward, W.W., Prendergast, F.G., Cormier, M.J. Primary structure of the Aequoria victoria green-fluorescent protein. *Gene* **1992**; *111*: 229–233.

18 Ng, G.Y.K.,Trogadis, J., Stevens, J., Bouvier, M., O'Dowd, B.F., George, S.R. Agonist-induced desensitization of dopamine D1 receptor stimulated adenylyl cyclase activity is temporally and biochemically separated from D1 receptor internalization. *Proc Natl Acad Sci USA* **1995**; *92*: 10157–10161.

4

Importance of GPCR Dimerization for Function:
The Case of the Class C GPCRs

Laurent Prézeau, Cyril Goudet, Philippe Rondard, and Jean-Philippe Pin

4.1
Introduction

Most membrane receptors, including ligand-gated channels, tyrosine kinase receptors, cytokine receptors and guanylate cyclase receptors form oligomers. This was rapidly recognized as being crucial for the functioning of these receptors. In the case of ligand-gated channel receptors, association of three to five subunits is required to form an ion channel. In the case of receptors that have a single transmembrane domain, it was rapidly proposed that ligand binding in the extracellular domain induces receptor dimerization, allowing the associated intracellular effector domains to interact with each other and to be activated. More recent data arising from the determination of the 3D structure of the extracellular domains of such receptors with and without agonists revealed that they can even be constitutive dimers, the agonists stabilizing a specific active conformation of the dimer [1, 2].

For many years, G protein-coupled receptors (GPCRs) were therefore the only membrane receptors assumed to function as monomeric entities. These proteins have a large membrane core domain (HD, heptahelical domain) composed of seven transmembrane-spanning helices (TM) that are responsible, in most cases, for both ligand recognition and activation of the intracellular effectors, i.e. the heterotrimeric G-protein. Biochemical and biophysical data even confirmed that these proteins can oscillate between various conformations, the active conformations being stabilized by agonists, whereas the fully inactive conformations are stabilized by inverse agonists. Accordingly, there was no apparent need for oligomerization to explain GPCR activation of G-proteins. Moreover, both biochemical and biophysical data were consistent with monomeric rhodopsin being able to activate transducin [3]. However, it was difficult to explain the cooperativity phenomenon observed in ligand binding in many GPCRs, suggesting that these proteins could form oligomers. This was then firmly demonstrated for most GPCRs using both co-immunoprecipitation experiments and energy transfer technologies [4]. Within the last few years, a number of studies indicated that this phenomenon is involved in trafficking the receptor to and from the plasma membrane, and in spe-

cific cross-talk between receptor subtypes [5]. However, the precise role and importance of GPCR oligomerization in the activation process remains unknown. Understanding the functional significance of this process is crucial for drug development. Indeed, if GPCRs really need to form dimers to activate G-protein, it is to be expected that cross-talk between the binding sites may open new possibilities for the development of drugs which specifically target GPCR dimers. In addition, this may also help unravel the functional role of many orphan receptors that may possibly function only when associated with another GPCR.

Five main classes of GPCRs have been defined in mammals based on sequence similarity [6–8]. Class A comprises the large number of rhodopsin-like receptors, whereas the secretin-like and metabotropic glutamate (mGlu)-like receptors are members of classes B and C, respectively. The two other classes correspond to the Frizzled-like receptors and a subgroup of pheromone receptors, respectively. Among these various GPCR types, those of the class C are constitutive dimers, both subunits being linked in most cases by a disulfide bridge. These receptors therefore represent an appropriate model for the study of the functional relevance of GPCR dimerization. Class C GPCRs include the receptors for the main neurotransmitters, glutamate and GABA, as well as receptors activated by arginine, extracellular Ca^{2+}, some pheromones, and sweet and umami taste compounds [9].

In this chapter, we will summarize our current knowledge regarding the functioning of class C GPCRs, and illustrate how allosteric interactions between the subunits play a fundamental role in their activation. Although these findings may well be specific for this class of GPCRs, they suggest new possibilities concerning the functioning of receptors from the other classes.

4.2
Class C GPCRs are Multidomain Proteins

Members of class C GPCRs can sense a large variety of molecules, ranging from amino acids (mGlu1–8, GABAB receptors, and OR5.24), to calcium ions (CaS, calcium-sensing receptors), and volatile pheromones or soluble taste molecules (V2R, T1R2/T1R3 sugar and T1R1/T1R3 umami taste receptors, and OR5.24 odorant receptor) [9]. These receptors share a common complex structure (Fig. 4.1). First, as with all GPCRs, they possess a HD which couples to the G proteins. This HD domain is formed by seven TMs (TM1 to 7) connected by extracellular and intracellular loops (e and i loops, respectively), and terminates with an intracellular C-terminal tail (C-tail). But second, the original feature of these receptors is the presence of a large extracellular domain (ECD), divided in two regions, a large Venus Fly-Trap module (VFT) that binds the endogenous ligand, and in most of the class C receptors a smaller Cysteine Rich Domain (CRD) located between the VFT and the HD [9]. The modular composition of these receptors together with their dimerized nature, make their functioning more complex than that of most of the other GPCRs.

We will see how these receptors function, i. e. how each module plays its own part in producing a surprising symphony of action through allosteric interaction.
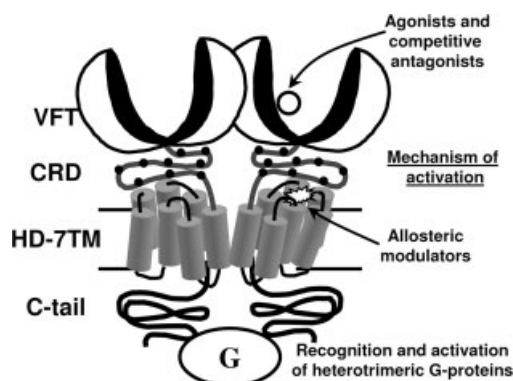
Fig. 4.1 Schematic representation of the main domains of class C GPCRs. Class C GPCRs are composed of several main structural domains, the Venus Flytrap domain (FVT) where agonists and competitive antagonists bind, the cysteine rich domain (CRD) that interconnects the VFT to the third domain, i. e. the hepta-helical domain (HD) similar to rhodopsin-like GPCRs, and a variable C-tail. The ligand binding site and the site of action of the allosteric compounds are indicated in the VFT and in the HD, respectively. G = G protein.

4.2.1
The VFT

Among the GPCRs, the class C receptors (except orphan members related to the protein Boss [9]) have a large VFT of around 500 residues.

Very early on, the VFT was shown to contain the endogenous ligand binding site, either by mutagenesis or by expressing the VFT alone in a soluble form able to bind the ligands [10–14]. One of the most important findings in the field was the observation made by O'Hara in 1993 who first reported a sequence homology between the VFT and some periplasmic bacterial proteins (PBP), suggesting that they have a similar structure [15]. This allowed the generation of 3D models of mGlu, CaS and GABAB receptor VFT which were validated by mutagenesis studies (e. g. [16–19]). Thus, according to the functioning of the PBP, it was assumed that the VFT was folded into two lobes separated by a cleft in which the ligand binds (Fig. 4.1), a structure which resembles a fly-trapping carnivore plant. As is the case with PBPs, it was thought that the lobes exist in two different states, open and closed, driven by an equilibrium constant between the two forms [15]. This was confirmed by the determination of the crystal structure of the VFT of mGlu1 [20, 21].

Indeed, the VFT of mGlu1 was crystallized in several open and close states [20, 21]. Without any ligand or in the presence of an antagonist, the VFT was open, but both open and closed forms were also obtained in the presence of glutamate. This led to the conclusion that an agonist stabilizes the closed form and an antagonist the open form, leading to a ligand affinity that does not only depend on its binding in the cleft, but also on the equilibrium constant controlling the open and closed states [22].

The glutamate amino-acid moiety interacts with the conserved residues serine, Ser165, and threonine, Thr188, in mGlu1, and the distal carboxylic group interacts

with the conserved lysine, Lys409, and arginine, Arg78, in mGlu1, the four of them being located in the lobe I. However, glutamate also interacts with lobe II residues, especially the tyrosine Tyr236, aspartate, Asp318, and glutamate, Glu292, in mGlu1, which are important for stabilizing the closed state upon agonist binding. This pocket is conserved in mGlu receptors, and is similar to that of the PBPs [9]. The residues involved in the binding of the amino acid moiety were conserved in receptors known to bind amino acids, such as the CaS and some of the taste receptor subunits and in the fish olfactory receptor OR5.24 and its mammalian ortholog GPRC6A, or partly conserved, as in the GABAB1 receptor [9].

4.2.2
The CRD

The CRD is a 70-residue domain, containing nine conserved cysteines, and is present in mGlu, CaS, taste and pheromone receptors, but not in GABAB receptors. A recent publication proposed a folding of the CRD of mGlu1 receptor with 4 β strands and three disulfide bridges [23] but no data supporting this model have been published as yet.

The CRD seems to be necessary for receptor activation, as the CaS receptor is not functional after deletion of its CRD [24]. The human CaS receptor in which the CRD is replaced by the CRD of the fish FuGu CaS receptor is functional, however. The replacement of the CRD by the CRD of mGlu1 receptor led to a receptor still activated by Ca but with greatly reduced function. Moreover, a chimeric receptor containing the VFT of CaS and the HD of mGlu1 needs a CRD from either the CaS or the mGlu1 receptor to become functional. The CRD could be involved in the transduction of the signal between the VFT and the HD. But although the VFT and the CRD have few free cysteine residues which are not involved in intradomain disulfide bonds, both domains are probably not linked by disulfide bond [25].

It is noteworthy that several natural mutations leading to disease have been found in the CRD of the CaS receptor [23, 26]. The mutation C582Y is thus associated with severe neonatal primary hyperparathyroidism, and mutations G549R and G557E in the human CaS receptor correlate with familial hypocalciuric hypercalcemia.

Interestingly, in the heterodimeric T1R2/T1R3 sweet taste receptor, the CRD of T1R3 is required to recognize the specific sweet taste of the protein Brazzein, a naturally-occurring sweet-tasting plant protein, and is also required but to a lesser extent to recognize the sweet taste of another plant protein, Monellin [27]. Thus, the CRD may participate in the detection of some ligands.

In any case, the precise function of this region is not yet known. Nevertheless, as the CRD is located between the VFT and the HD, it is tempting to hypothesize that it plays a role in the activation process [22] or that it could be involved in the interaction with regulatory proteins or in dimerization. The discrepancy with the GABAB receptor, which does not possess a CRD but is perfectly functional, has yet to be elucidated.

4.2.3
The HD

The first function of the HD is to couple to G proteins via the i loops and an H8 helix located just after the TM7. The H8 region in rhodopsin is an amphipatic helix parallel to the plasma membrane that regulates the coupling efficiency. Indeed, as in rhodopsin-like receptors, mutations in i2, i3 loops or H8 impaired the coupling to G proteins in mGlu, Cas and GABAB receptors [28–31]. It has also been shown that the protein coupling is regulated by phosphorylation of the loops or the C-tail [32, 33]. Moreover, it should be noted that the i3 loop of mGlu6 can functionally replace the i2 loop in rhodopsin [34], supporting the assertion that the loops in class C receptors and rhodopsin-like receptors play the same role, i.e. to couple with the G proteins.

To achieve G protein stimulation, the HD should be turned on, but the mechanism of activation is still unknown. The activation of class A and B receptors is better understood, but the HD of class C receptors does not share significant sequence homology with the HD of class A and B GPCRs. Nevertheless, several clues suggest that they fold and function in a similar way.

First, GPCRs from the different classes couple to the same G proteins, and the regions of the receptors dedicated to G protein coupling are the same [35].

Second, few residues are nevertheless conserved between receptors of classes A, B and C, consistent with the hypothesis that these HDs originated from a common ancestor. The two cysteine residues in the extracellular face of TM3 and in the e2 loop known to form a disulfide bond in class A GPCRs, are conserved in the class C GPCRs. Moreover, the tryptophan in TM6, important for the activation mechanism of class A receptors, is also found in most class C members. An arginine residue at the intracellular face of TM3 may correspond to the arginine of the class A DRY (aspartate–arginine–tyrosine) motif, known to be involved in activation of class A GPCRs (for review see [9]).

The third argument comes from the analysis of the hydrophobicity of TM3. Actually, TM3 in mGlu receptors is weakly hydrophobic, suggesting that as in the rhodopsin-like receptor, TM3 may be central in the organization of the HD structure of mGlu receptors.

Fourth, it is also worth noting that 3D models of the HD of mGlu receptors, based on the crystal structure of rhodopsin [36] correlate well with mutagenesis studies. For example, the binding pocket of the allosteric modulators in the HD is compatible with the position of the TMs defined by such 3D models [37–40].

As seen above, the endogenous binding site is located in the VFT. But in the last few years, an increasing number of molecules have been described which act directly in the HD of mGlu, CaS, sweet taste or GABAB receptors [41, 42]. Indeed, negative and positive allosteric modulators inhibit or facilitate the action of classical agonists respectively, by direct binding to a pocket delineated by TM3, 5, 6 and 7, that may correspond to the binding pocket of the retinal in rhodopsin [41]. This pocket could thus be a fossil binding pocket of the common ancestral HD.

Moreover, in contrast to the glutamate ligand binding pocket that is highly conserved in the VFT of the different mGlu receptors, the site of action in the HD allows a greater selectivity of action. For example, MPEP and CPCCOEt or BAY367620 specifically inhibit mGlu5 and mGlu1 respectively, while the competitive antagonists usually inhibit both of them. The positive allosteric modulators for mGlu, GABAB, CaS, or taste receptors, generally have little effect by themselves, but serve to increase the action of the agonists [41]. This property makes them interesting therapeutic molecules, as they will probably induce fewer side-effects, such as desensitization of receptors, because they act only in the presence of the agonist.

More surprisingly, agonists have also been described to act in the HD, especially for the dimeric sweet taste receptor T1R2/T1R3, as some sweeteners such as cyclamate activate the HD of the sweet taste receptor [43]; since these agonists act on the T1R3 subunit which is common to sweet (T1R2/T1R3) and umami (T1R1/T1R3) receptors, cyclamate is also an allosteric regulator of umami taste. Allosteric compounds with agonist activity have also been described to act in the HD of GABAB receptor [44].

4.2.4
C-Tail

We consider the C-tail to start after the H8 loop described above. The C-tail is very variable in size (between 10 to 360 residues long) due to receptor diversity and alternative splicing.

In addition to the fact that the C-tail possesses phosphorylation sites serviced by kinases which regulate the activity of the receptors, the C-tail usually contains numerous identified regions (acidic, serine-rich or proline-rich regions), and many interaction sites or motifs that allow association with regulating (calmodulin, G protein receptor kinases (GRK), etc.), scaffolding (Homer, 14–3-3, PICK1, etc.) or signal transduction (protein kinase C (PKC), α, β, γ subunits of G proteins, etc.) proteins (for review see [45]). These interacting proteins may control about every step in the life of a receptor: targeting to the correct membrane compartment, its activity, its association with other receptors, its desensitization and internalization, etc.

The modular structure of these receptors makes their functioning quite complex. But their dynamic mechanism is even more complex, as these receptors are dimeric functional entities.

4.3
Class C GPCRs are Constitutive Dimers

Our view of the structural complexity of class C GPCRs further increased when Romano et al. reported that mGlu5 receptors exist as constitutive homodimers, both in transfected cells and in native brain tissue [46]. Indeed, data indicated that

each subunit is likely linked by a disulfide bridge at the level of the VFT. A disulfide covalent linkage between the subunits was later further documented for most mGlu receptors, as well as for the CaS receptor [47–49]. The Cys involved was identified in mGlu1, mGlu5 and the CaS receptors [47–50]. It is located in a large loop in lobe I of the VFT and corresponds to a residue conserved in all class C GPCRs possessing a VFT (i.e. except the GABAB subunits). Although this disulfide bridge increases the stability of the protein and prevents any possible dissociation of the dimer, it is not required for dimer formation [49, 50]. When produced in soluble form, and isolated from the rest of the protein, the VFTs form dimers [12, 14, 49], and this was definitively confirmed by the crystallization of the VFT of mGlu1. Indeed, the solvated structure revealed a large hydrophobic dimer interface at the level of lobe I of the VFT [20].

The importance of class C dimer formation for function was strengthened when it was discovered that the GABAB receptor needs two different subunits, GABAB1 and GABAB2, to form a functional receptor [51–54]. Neither subunit induced any response when expressed alone; it emerged that each subunit plays a specific role in the dimer. Whereas GABAB1 is responsible for GABA binding, the GABAB2 subunit plays a critical role in G protein coupling [18, 30, 55–57]. The importance of heterodimer formation has now been confirmed *in vivo* since both GABAB1 and GABAB2 knock-out mice share an identical phenotype, and classical GABAB responses can no longer be detected in either case [58–61]. Moreover, a quality control system has been selected during evolution to insure that only the heteromeric entity reaches the cell surface. Indeed, the GABAB1 subunit possesses an intracellular retention signal RSR (arginine–serine–arginine) in its C-terminal tail, that is masked when this subunit is associated with GABAB2 [62, 63].

More recently, the class C taste receptors were also shown to function as heterodimeric entities, the sweet and umami receptors being composed of T1R2 and T1R3, and T1R1 and T1R3 subunits, respectively [64, 65]. Although the T1R3 subunit is common to both receptor types, it appears that the associated subunit T1R2 and T1R1 are both responsible for ligand-specific activity and G-protein coupling [43].

Taken together, these observations demonstrate that class C GPCRs are large protein complexes composed of two subunits each being made up of four identified domains, the VFT, the CRD, the HD and the C-tail. These data also indicate that the dimeric nature of these receptors is likely required for function. This raises a number of important issues, one being how does agonist binding in the VFT trigger receptor activation?

4.4
Agonists Activate Class C GPCRs by Stabilizing the Closed State of the VFT

As described above, and in agreement with what was reported for PBPs, the mGlu1 VFT can adopt either an open conformation observed without ligand, or with bound agonist or antagonist, or a closed conformation observed with bound

agonist only [20, 21]. It was therefore soon assumed that agonists act by stabilizing the closed state of the VFT, whereas antagonists would prevent closure. However, these specific conformations of the VFT were only observed with the soluble form and therefore in the absence of the rest of the receptor.

To examine whether VFT closure was required for receptor activation in wild-type receptors, Bessis et al. examined the mechanism of action of two mGlu receptor antagonists that differ only slightly from agonists. The two antagonists were alpha-methyl-aminophosphonobutyrate (MAP4), the alpha-methyl derivative of the potent group III mGlu agonist L-aminophosphonobutyrate (L-AP4), and (1R,3S,4R)1-aminocyclopentane-1,3,4-tricarboxylate (ACPT-II), a stereoisomer of the potent agonists ACPT-I and (+)-ACPT-III [16]. When docked into a closed form model of the mGlu8 VFT, steric and ionic hindrance due to aspartate D309 and tyrosine Y227 were observed, respectively, consistent with the idea that these two antagonists cannot fit into the closed form pocket, and therefore do not allow VFT closure [16]. After mutation of D309 or Y227 to alanine residues to remove the hindrance, both antagonists were converted into full agonists, demonstrating that the ability of a ligand to bind in the closed form of the VFT is required for agonist activity.

The data reported by Bessis et al. demonstrated that stabilization of the closed VFT is required for mGlu receptor activation. However, this poses the question of whether this is the only effect of agonists. In other words, is the closure by itself sufficient to activate such multidomain and dimeric receptors? To answer this question, Kniazeff et al. inserted two cysteines on each side of the cleft of the GABAB1 VFT such that a disulfide bridge is expected to lock this domain in a closed state [66]. The mutated receptor displayed high constitutive activity that was suppressed after treatment with reducing agents such as DTT. Moreover, the activity of the disulfide bridged receptors could not be inhibited by the antagonist CGP64213, and indeed these antagonists did not bind to the cross-linked binding domain, consistent with the VFT being in a closed form [66].

Taken together, these data neatly demonstrate that the agonists act by stabilizing the closed form of the VFT binding domain, whereas competitive antagonists prevent such closure.

4.5
Dimeric Functioning of the Dimer of VFTs

Although VFT closure is associated with a major change in conformation of the VFT dimer, the question of how the closure of the VFT led to the activation of the HD still remained. Again, a major finding that helped answer this important question came from the resolution of the crystal structure of the mGlu1 VFT. As already mentioned, these studies confirmed that the VFTs assemble into dimers but, most importantly, also revealed a major change in the general conformation of the dimer in the presence of a bound agonist [20, 21].

In the absence of agonist, both VFTs are in the open state and interact at the level of a hydrophobic surface of their lobe I (Fig. 4.2 a) [20]. The relative orienta-

tion of the VFTs is such that there is no contact between their respective lobes II, and their C-termini are far apart (88 Å). Because the same structure has been obtained with a bound antagonist [21] (Fig. 4.2 a), it likely corresponds to the resting state of the receptor. Accordingly, this conformation of the VFT dimers of has been named Roo, for resting orientation with both VFTs in the open state.

In contrast, in the presence of glutamate, a new relative orientation known as active, is observed, in which the lobe IIs contact each other such that their C-termini are closer (63 Å) (Fig. 4.2 a). In the first published structure of the dimer of mGlu1 VFTs with bound glutamate, only one VFT is in the closed state, whereas the other is still in the open state [20]. This conformation of the VFT dimers has been named Aco. A second conformation has then been reported and called Acc, in which both VFTs are still in the active orientation, but both are in the closed state (Fig. 4.2 a) [21].

A detailed analysis of the lobe II interface provides a rationale for the agonist-induced change in the relative orientation of the VFTs in the dimer, from the R to the A orientation. Indeed, if none of the VFTs are closed, an ionic repulsion occurs at the level of the interface between the two lobe IIs, a basic residue from one VFT facing the same basic residue from the other, and a series of acidic residues also facing each other, making this A orientation of the VFT dimers highly unstable (see the hypothetical Aoo interface in Fig. 4.2 b). In contrast, if one VFT is in the closed state and the other still open (Aco), the lobe II interface is such that the basic residue now faces acidic residues, making this orientation stable (Fig. 4.2 b). In the symmetric Acc conformation, a new lobe II interface is observed, in which the acidic residues face each other, but can be stabilized by a gadolinium cation Gd^{3+} (Fig. 4.2 b). Accordingly, it was proposed that this Acc conformation can only be obtained *in vivo* if a cation such as Ca^{2+} or Mg^{2+} can bind to the receptor at the level of the lobe II interface. In agreement with this possibility, Ca^{2+} ions have been reported to enhance mGlu receptor function [67].

Taken together these structural observations provide a neat hypothesis for the dimeric functioning of class C GPCRs. But are these observations made with the soluble form of the VFTs valid for the full-length receptor? Several sets of data obtained from both the mGlu and GABAB receptors confirm that the activity of theses receptors is dependent upon the dimeric nature of the VFT. In the case of mGlu receptors, the mutation of residues at the lobe I interface suppresses agonist activation of the receptor [68]. In the case of the heterodimeric GABAB receptor, direct interaction between the VFTs of GABAB1 and GABAB2 has been reported [69], and the correct association of these two VFTs is required for GABA activation of the receptor. Indeed, a GABAB receptor combination in which both VFTs are of the GABAB1 type is not functional despite agonist binding to the receptor [70, 71].
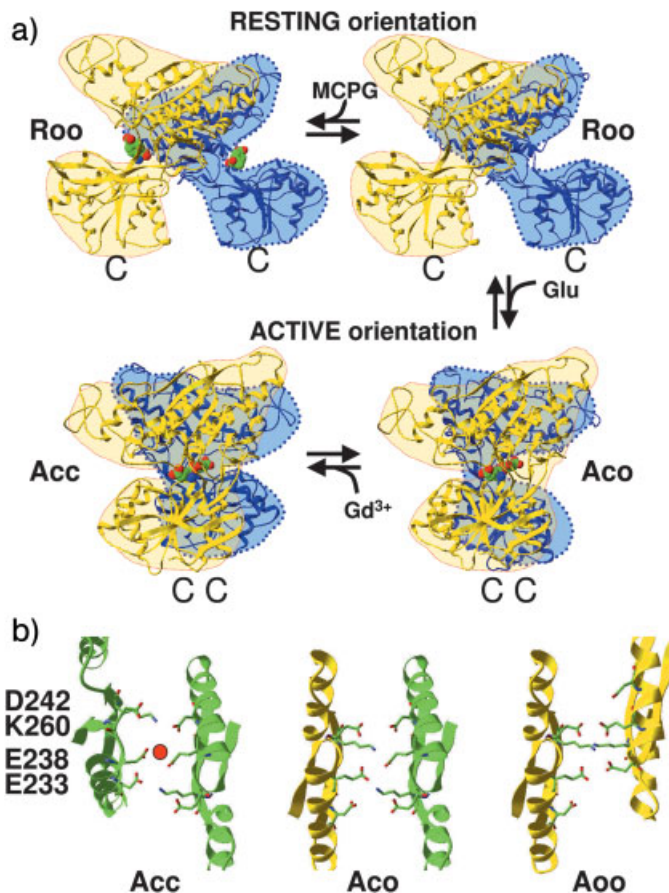
Fig. 4.2 General mechanism of activation of dimeric class C GPCRs. (a) Ribbon view of the crystal structure of the resting Roo (left, pdb accession number 1EWT) and fully active Acc (right, pdb accession number 1ISR) state of the mGlu1 VFT dimer, and apposition of two rhodopsin structures. The yellow subunit is in the front, while the blue subunit is in the back. Note the difference in the relative orientation of the two VFTs probably leading to a different mode of association between the two HDs within the dimer. (b) Ribbon view showing the interaction face between both lobe IIs of the VFT dimer in the different conformations. In an Aoo conformation, the interaction is not possible because of the repulsive action of the charged residues. Only the Aco and Acc conformations are accepted; the latter being stabilized in the presence of a positive ion neutralizing negative residues.

4.5.1
Agonist Stoichiometry: Symmetry or Asymmetry?

The essential nature of the dimeric form to the functioning of class C GPCRs raises the question of whether a single agonist per dimer is sufficient to activate the receptor or whether two agonists per dimer are required. The case of the

GABAB receptor offered a unique opportunity to clarify this issue since this receptor is an obligatory heterodimer. As described above, GABA has been shown to bind to the GABAB1 VFT [18, 55, 72], and the closure of this domain is sufficient to fully activate the receptor [66]. Moreover, a detailed analysis of the GABAB2 VFT revealed that this domain is unlikely to bind any natural ligand [66], further demonstrating that a single agonist is sufficient to activate the heterodimeric GABAB receptor. This conclusion is also consistent with the action of sweet compounds and glutamate (umami) on the taste heterodimeric receptors T1R2/T1R3 and T1R1/T1R3, respectively [64, 65, 73]. The functional analysis of mutant and chimeric T1R subunits indicates that glutamate binds to the T1R1 VFT while many sweet compounds interact with the VFT of T1R2 [43].

In the case of the homodimeric mGlu receptors, a single agonist is also sufficient to activate the receptor [74]. This was demonstrated by using the quality control system of the GABAB receptor, which allowed the dimeric mGlu receptors composed of a wild-type subunit and a mutant subunit which was unresponsive to glutamate, to target the cell surface. This is consistent with the structural data showing that a single closed VFT is sufficient to achieve the active orientation of the VFT dimers (the Aco conformation described above). However, the receptor was only partially active in the presence of a single agonist, and full activity was only achieved after the binding of a second agonist [74]. Further functional analyses of dimeric receptors composed of two defined subunits indicated that the receptor is only partially active when a single VFT is closed (Aco), and that full activity is only attained when both VFTs are in the closed state (Acc). This highlights the symmetric nature of the functioning of mGlu VFT dimers.

Taken together, these data indicate that a change in the relative orientation of the VFTs resulting from the closure of at least one of both domains is associated with receptor activation.

4.6
The Heptahelical Domain, the Target of Positive and Negative Allosteric Modulators, Behaves in a Manner Similar to Rhodopsin-like Class A GPCRs

The specific structural arrangement of the class C GPCRs questions whether or not the heptahelical domain of these receptors behaves like the other GPCRs, such as rhodopsin. As mentioned in Chapter 1, the HD of class C GPCRs shares a very low sequence similarity with the other receptors. However, a number of arguments suggest that both domains share a similar structure. The characterization of the mechanism of action of class C allosteric modulators helped to demonstrate that in accordance with the other rhodopsin-like GPCRs, the class C HD can exist in various conformational states from completely inactive to fully active.

As mentioned earlier, two types of allosteric modulators have been identified. The first were the non-competitive antagonists that were soon shown to act as inverse agonists [40, 75]. The second type corresponds to the positive allosteric modulators. Such compounds, first discovered for the CaS receptor, are either devoid

of agonist activity [40, 41, 76–83] or display a low partial agonist activity [44, 84]. However, they augment the potency and, in most cases, also the efficacy of agonists. So far, all these compounds had been shown to bind directly to the HD, in a cavity corresponding to the retinal binding pocket of opsins [38–40, 82, 85, 86]. Therefore such compounds are potential tools which may further our understanding of the activation mechanisms of the HD of class C GPCRs.

When they were discovered, it was assumed that negative allosteric modulators acted to stabilize the HD of the receptor in a resting conformation [38–40, 75], and that positive modulators would stabilize the active conformation [82]. This has been confirmed by recent studies showing that the HD alone of class C GPCRs expressed functions similar to the rhodopsin-like GPCRs [44, 83, 87]. In fact, the HD of mGlu5 when separated from its ECD is functional and displays the same constitutive activity as the wild-type receptor. This constitutive activity can be inhibited by inverse agonist/negative allosteric modulators. Moreover, the HD of mGlu5 or GABAB2 can be activated by positive allosteric modulators acting as full agonists [44, 83], as does the combination of the CaS positive modulator NPS568 and Ca^{2+} [87]. Thus, the HD is able function by itself, but its activation is controlled by the dimer of VFTs.

What is the explanation for the fact that positive allosteric modulators which directly activate the isolated HD are generally devoid of intrinsic agonist activity in the full length receptor? Recent studies suggest that GPCRs can oscillate between three main conformations: a resting "ground" state Rg which corresponds to a fully inactive state stabilized by an inverse agonist, a resting state R in which the receptor is able to weakly activate G-proteins, and an active state R* stabilized by agonist binding [88, 89]. By analogy, we proposed that the HD of class C receptors can also oscillate between three similar states, the totally inactive HDg, the partially inactive HD and the fully active HD* (Fig. 4.3). In the absence of agonist, the dimer of ECD is in its open resting state Roo and prevents the transition between HD and HD*. Depending on the receptor and its equilibrium constant, the
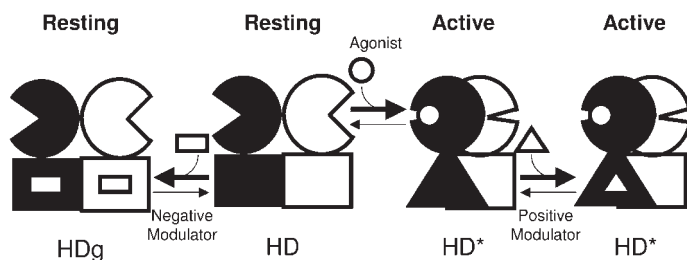
Fig. 4.3 Schematic representation of the putative mechanisms of action of negative and positive allosteric modulators. In the absence of an agonist, the receptor is in a resting state (Roo-HD), and switches to a partially active state upon binding of a first agonist (Aco-HD*), and to a fully active state upon binding of a second agonist (Acc-HD*). Binding of an inverse agonist in the HD stabilizes the fully inactive ground state of the receptor, whereas binding of a positive allosteric modulator further stabilizes the fully active state of the agonist bound dimer.

HD can oscillate between HDg and HD and a constitutive activity can be observed. Negative allosteric modulators stabilizing the HDg state act as inverse agonists. The transition between HD and HD* is not possible when the VFTs are in their resting conformation thus preventing the positive allosteric modulators to act as agonists. However, upon agonist binding, the dimer of VFTs reaches its active state Acc and triggers the transition from HD to HD*, the latter being further stabilized by positive allosteric modulators. The dimer of HD is dependent on the state of the dimer of VFTs and on the constant that rules their coupling. Moreover, the fact that the HD behaves in a different manner to the wild-type receptor when expressed alone, suggests that the dimer of VFTs has a constitutive inhibitory action on the dimer of HDs. The precise mechanism has yet to be elucidated.

4.7
Allosteric Coupling Between the Extracellular and Heptahelical Domains within the Dimer

4.7.1
Molecular Determinants of the Coupling Between the VFT and the HD

The VFT and the HD are connected through a CRD within most class C GPCRs, with the exception of GABAB. Although the specific role of the CRD in the activation process is not known, it seems to be a necessary requirement for the allosteric coupling between the VFT and the HD of either mGlu or CaS receptors to take place [24] as discussed above. Although a tridimensional model of the CRD was recently proposed [23], the exact positioning of this domain relative to the VFT and the HD is unknown. It could be physically located either between the VFT and the HD, or alternatively, on the side of the VFT and HD domains. Eventually, it was shown that the CRD is not linked to the VFT domain by a disulfide bridge [25].

In the GABAB receptor subunits, an ~ 40-amino acid long peptide linker connects the VFT with the HD. This peptide linker can be altered, and even be made longer without affecting receptor function [56, 71], suggesting that it is not crucial to the allosteric coupling between the VFT and the HD in these GABAB receptor subunits. Also, it indicates that direct contact between the VFT and the HD underlies ligand-induced signal transduction within this receptor.

Evidence for a direct contact between the VFT and the HD has been obtained by measuring agonist binding affinity. Negative allostery within GABAB1 expressed alone is due to a direct interaction between the HD and the VFT of this subunit [69]. Indeed, the HD constrains the VFT in a low agonist affinity state, as demonstrated by the increased agonist affinity when the HD is deleted [13, 69]. This negative allostery is controlled by the GABAB2 VFT. Such a negative effect of the HD on agonist affinity has also been reported for other class C GPCRs, such as the mGlu4 and mGlu8 receptors [11, 14].

4.7.2
Cis- and *Trans*-activation Can Exist within Class C GPCRs

Since as described above a single agonist can activate class C GPCRs, is the bound VFT able to activate the HD of the same subunit (*cis*-activation), or that of the other subunit (*trans*-activation), or both? In the wild-type heterodimeric GABAB receptor, it seems that only *trans*-activation occurs since GABA binding in the GABAB1 VFT leads to the activation of the GABAB2 HD. However, normal functioning of the GABAB receptor was observed when the VFTs were swapped between the two subunits [70, 71]. In that case, the agonist binding domain (the GABAB1 VFT) and the G-protein activating domain (the GABAB2 HD) are fused in the same subunit, indicating that *cis*-activation can also occur in class C GPCRs.

In agreement with the above conclusion, both *cis*- and *trans*-activation occur in homodimeric mGlu receptors. As described above, by manipulating each subunit in a receptor dimer, it was shown that the monoliganded dimer of VFTs in the Aco conformation lead to partial activity, whereas the Acc conformation with two bound agonists lead to full activity of the receptor [74]. By examining the effect of a point mutation known to prevent G-protein activation in the i3 loop of either HD, it was shown that both the Aco and Acc conformations of the VFT dimer activate either one or the other HD in the receptor dimer [74].

This observation neatly demonstrates that the closed state of one VFT can activate either the HD of this subunit or that of the associated subunit. This is consistent with the change in the relative orientation of the VFTs being the "trigger" for HD activation, rather than the VFT closure *per se*.

The mechanism of activation of the dimer of HD is not yet known. A first study recently elucidated some interesting information. Tateyama et al. recently reported data consistent with a change in the relative position of the HDs in a mGlu1 dimer [90]. Indeed, by inserting fluorescent probes (CFP and YFP) which allow the transfer of fluorescence energy (FRET), into the i loops of mGlu1, they showed that agonists modified the FRET between the two protomers, as the FRET between the two i2 loops increased and that between the two i1 loops decreased. In contrast, the FRET between two probes inserted into two loops of the same protomer to monitor the internal conformational changes, was not changed. This does not rule out internal HD modification of conformation, as these changes may not be monitored by the technical strategy used.

These data suggest that two events occur at the level of the HD dimer during the activation process: (1) a change in the relative position of the HDs within the dimer, and (2) as observed in the other GPCRs, a change in the conformation of the HDs. But does this change in conformation occur in one or both HDs?

4.8
Asymmetric Functioning of the HD Dimer

In the light of recent data on GPCR oligomerization, the classical model of G-protein activation which was based on the idea that a monomeric receptor is activated by one ligand and activates a single heterotrimeric G-protein, has been re-examined.

First, based on structural and crystallographic studies, two receptor molecules seem to be required to offer the presumed interface of contact with a single G-protein [36, 91, 92]. While several authors propose that a dimer of GPCR is necessary to activate one G-protein [93–95], it is puzzling to note that transducin can be activated by either dimeric or monomeric rhodopsin [3, 96] and that in heterodimeric GABAB and taste receptors only one subunit appears to be critical for G protein activation [30, 43, 64, 70]. This implies that each HD in a dimer does not play the same role, introducing asymmetry within the dimer. Moreover, although the symmetrical binding of two agonists per mGlu dimer is required for full activation, the binding of a single agonist is sufficient to obtain measurable activity [74]. This raises the question of the symmetric functioning of both subunits in the homodimeric receptor activation process from ligand binding to G protein activation.

In two series of experiments, we examined whether one or both HDs able to activate G-proteins were necessary for full receptor activity. In the first series, we examined the coupling efficacy of a GABAB receptor containing two GABAB2 HDs. In such a receptor containing an HD homodimer, the mutation of either one in the i2 or i3 loop (mutations known to prevent G-protein activation), decreased the G-protein coupling efficacy of the dimer by a factor of 2. Only when both subunits were mutated was the G-protein activation suppressed. Similar data were obtained with both mGlu5 [74] and mGlu1 [97] receptors. In that case, the quality control system of the GABAB receptor was used to largely increase the proportion of receptor dimers composed of two defined subunits at the cell surface, one with the wild-type HD and the other with a mutation in its i3 loop.

Two hypotheses could explain the fact that a receptor dimer bearing a mutation in a single subunit is less efficient at activating G-proteins than the wild-type homodimer. The first is that each HD is able to bind and activate G-proteins independently of the other. However, based on crystallographic and structural studies, the binding of two heterotrimeric G-proteins under a dimer of GPCRs seems difficult to achieve [91, 92]. The second hypothesis is that only one HD is turned on at a time.

Studying the stoichiometry of negative allosteric modulators in an mGlu1 receptor dimer provided results supporting this second hypothesis [98]. Indeed, in a receptor dimer possessing a single subunit sensitive to the negative modulator, no effect of such compound was observed. This observation indicates that in contrast to the mutation of the i2 or i3 loop, stabilizing one HD within the dimer in its inactive conformation does not affect G-protein coupling efficacy of the dimer. However, a total loss of coupling is observed if one HD is impaired in its ability to activate G-proteins while the other HD is stabilized in its inactive state by a negative

allosteric modulator. In contrast, if the negative allosteric modulator binds to the HD which is unable to activate the G-protein then it acts as an enhancer, potentiating the effect of agonists by driving the G-protein activation toward the other HD. All these results are in agreement with a model in which at any one time only a single HD is turned on per dimer [98].

The asymmetric activation process of the HD dimer is quite surprising since the activation of the dimer of VFTs is apparently symmetric. Indeed, the fully active state of the dimer is only reached when both VFTs have bound glutamate and are in their closed conformation [74]. If one considers that a dimer of GPCRs interacts with a single heterotrimeric G-protein then it could be hypothesized that the interaction with the G-protein brings about an external constraint that prevents both HDs from behaving in a similar manner.

4.9
Conclusion

Although being among the last GPCRs to be discovered, and despite their complex structural arrangement, much information has been obtained within the last 10 years concerning the activation process of class C GPCRs. This information is consistent with the following model of class C GPCR activation (Fig. 4.3). Binding of an agonist in the VFT of one subunit within the dimer allows the closure of this domain, and, as a consequence leads to the stabilization of a new relative orientation of the two VFTs in the dimer, and the stabilization of the second VFT in a closed form. This large change in the relative orientation of the VFT dimer probably has two main consequences for the HD dimer: (1) a change in their relative position; and (2) a change in conformation in only one protomer.

Still, a number of important issues remain to be solved: how does the active orientation of the VFTs lead to a change in the conformation of one HD? Is it the consequence of the direct interaction between the ECD (either the VFT or the CRD) and the HD, or is it the consequence of the change in the relative position of the HDs in the dimer? Why is only one HD activated, is it really due to the interaction of the receptor with the asymmetric heterotrimeric G-protein? What is the interface of the two subunits at the level of the HD? What is the specific role of the CRD?

Whatever the answers to these questions, this complex activation process of class C GPCRs offers a number of interesting possibilities for the control of their activity with synthetic compounds, including those acting at the already identified sites in the VFT and the HD, but also likely at any possible site at the interface between the subunits.

These observations also provide important new information that may be useful for the understanding of the activation process of other GPCRs. One piece of information that is of major importance is that although class C GPCRs are dimers, at any one time only a single HD is converted into the active state. This is consistent with data indicating that a monomeric rhodopsin can activate transducin [3],

but raised the question of the significance of GPCR dimerization in G-protein activation. This is puzzling when considering the recent observation that a single heterotrimeric G-protein may interact with a GPCR dimer, the ras-like domain of the α-subunit interacting with one protomer, while βγ and the N-terminal α helix of the α-subunit contact the other protomer [36, 94, 95]. This is also puzzling when considering the large number of GPCRs which have been described as dimers both in recombinant systems and native tissue, as revealed for example for rhodopsin in native disk membranes using atomic force microscopy [93]. In any case our observation does not contradict the involvement of GPCR dimers in G-protein activation, or indeed, that a single HD per dimer reaches the active state at any one time is consistent with a number of data obtained with many other GPCRs. One such example is the detection by the retina of a number of photons. Accordingly, in a rhodopsin dimer, a single retinal will be excited and converted into the activating trans-retinal, while that of the associated subunit in the dimer will likely remain in the *cis*-configuration corresponding to an inverse agonist. A second example is the case of the CCR2 : CCR5 heterodimer [99], in which agonist binding in one subunit inhibits agonist binding in the associated subunit, consistent with a single HD in a dimer being able to reach the active conformation. If this is proven to be the case for most GPCR dimers, and especially heterodimers, this will need to be taken into consideration for the process of drug development.

References

1 He, X.-L., D.-C. Chow, M.M. Martick, K.C. Garcia. Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. *Science* **2001**; *293*: 1657–1662.

2 Livnah, O., E.A. Stura, S.A. Middleton, D.L. Johnson, L.K. Jolliffe, I.A. Wilson. Crystallographic evidence for preformed dimers of erythropoietin receptor before ligand activation. *Science* **1999**, *283*: 987–990.

3 Jastrzebska, B., T. Maeda, L. Zhu, D. Fotiadis, S. Filipek, A. Engel, R.E. Stenkamp, K. Palczewski. Functional characterization of rhodopsin monomers and dimers in detergents. *J Biol Chem* **2004**; *279*: 54663–54675.

4 Bouvier, M. Oligomerization of G-protein-coupled transmitter receptors. *Nat Rev Neurosci* **2001**; *2*: 274–286.

5 Terrillon, S., M. Bouvier. Roles of G-protein-coupled receptor dimerization. *EMBO Rep* **2004**; *5*: 30–34.

6 Kolakowski, L.F. GCRDb: a G-protein-coupled receptor database. *Receptors Channels* **1994**; *2*: 1–7.

7 Fredriksson, R., M.C. Lagerstrom, L.G. Lundin, H.B. Schioth. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

8 Bockaert, J., J.-P. Pin. Molecular tinkering of G-protein coupled receptors: an evolutionary success. *EMBO J* **1999**; *18*: 1723–1729.

9 Pin, J.-P., T. Galvez, L. Prezeau. Evolution, structure and activation mechanism of family 3/C G-protein coupled receptors. *Pharmacol Ther* **2003**; *98*: 325–354.

10 Takahashi, K., K. Tsuchida, Y. Tanabe, M. Masu, S. Nakanishi. Role of the large extracellular domain of metabotropic glutamate receptors in agonist

selectivity determination. *J Biol Chem* **1993**; *268*: 19341–19345.

11 PELTEKOVA, V., G. HAN, N. SOLEYMANLOU, D.R. HAMPSON. Constraints on proper folding of the amino terminal domains of group III metabotropic glutamate receptors. *Brain Res Mol Brain Res* **2000**; *76*: 180–190.

12 OKAMOTO, T., N. SEKIYAMA, M. OTSU, Y. SHIMADA, A. SATO, S. NAKANISHI, H. JINGAMI. Expression and purification of the extracellular ligand binding region of metabotropic glutamate receptor subtype 1. *J Biol Chem* **1998**; *273*: 13089–13096.

13 MALITSCHEK, B., et al. The N-terminal domain of gamma-aminobutyric Acid(B) receptors is sufficient to specify agonist and antagonist binding. *Mol Pharmacol* **1999**; *56*: 448–454.

14 HAN, G., D.R. HAMPSON. Ligand binding to the amino terminal domain of the mGluR4 subtype of metabotropic glutamate receptor. *J Biol Chem* **1999**; *274*: 10008–10013.

15 O'HARA, P.J., et al. The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. *Neuron* **1993**; *11*: 41–52.

16 BESSIS, A.-S., P. RONDARD, F. GAVEN, I. BRABET, N. TRIBALLEAU, L. PRÉZEAU, F. ACHER, J.-P. PIN. Closure of the Venus Flytrap module of mGlu8 receptor and the activation process: insights from mutations converting antagonists into agonists. *Proc Natl Acad Sci USA* **2002**; *99*: 11097–11102.

17 COSTANTINO, G., R. PELLICCIARI. Homology modeling of metabotropic glutamate receptors. (mGluRs) structural motifs affecting binding modes and pharmacological profile of mGluR1 agonists and competitive antagonists. *J Med Chem* **1996**; *39*: 3998–4006.

18 GALVEZ, T., et al. Mutagenesis and modeling of the GABA-B receptor binding site suggest a Venus Fly-trap mechanism for ligand binding. *J Biol Chem* **1999**; *274*: 13362–13369.

19 HAMPSON, D.R., X.P. HUANG, R. PEKHLETSKI, V. PELTEKOVA, G. HORNBY, C. THOMSEN, H. THOGERSEN. Probing the ligand-binding domain of the mGluR4 subtype of metabotropic glutamate receptor (In Process Citation). *J Biol Chem* **1999**; *274*: 33488–33495.

20 KUNISHIMA, N., et al. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **2000**; *407*: 971–977.

21 TSUCHIYA, D., N. KUNISHIMA, N. KAMIYA, H. JINGAMI, K. MORIKAWA. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd^{3+}. *Proc Natl Acad Sci USA* **2002**; *99*: 2660–2665.

22 PARMENTIER, M.-L., L. PRÉZEAU, J. BOCKAERT, J.-P. PIN. A model for the functioning of family 3 GPCRs. *Trends Pharmacol Sci* **2002**; *23*: 268–274.

23 YU, L., S. LIANG, X. LIU, Q. HE, D.J. STUDHOLME, Q. WU. NCD3G: a novel nine-cysteine domain in family 3 GPCRs. *Trends Biochem Sci* **2004**; *29*: 458–461.

24 HU, J., O. HAUACHE and A.M. SPIEGEL. Human Ca^{2+} receptor cysteine-rich domain. Analysis of function of mutant and chimeric receptors. *J Biol Chem* **2000**; *275*: 16382–16389.

25 HU, J., G. REYES-CRUZ, P.K. GOLDSMITH, A.M. SPIEGEL. The Venus's-flytrap and cysteine-rich domains of the human Ca2+ receptor are not linked by disulfide bonds. *J Biol Chem* **2001**; *276*: 6901–6904.

26 HU, J., A.M. SPIEGEL. Naturally occurring mutations of the extracellular Ca^{2+}-sensing receptor: implications for its structure and function. *Trends Endocrinol Metab* **2003**; *14*: 282–288.

27 JIANG, P., Q. JI, Z. LIU, L.A. SNYDER, L.M.J. BENARD, R.F. MARGOLSKEE, M. MAX. The cysteine-rich region of T1R3 determines responses to intensely sweet proteins. *J Biol Chem* **2004**; *279*: 45068–45075.

28 CHANG, W., T.H. CHEN, S. PRATT, D. SHOBACK. Amino acids in the second and third intracellular loops of the parathyroid Ca^{2+}-sensing receptor mediate efficient coupling to phospholipase C. *J Biol Chem* **2000**; *275*: 19955–19963.

29 Francesconi, A., R.M. Duvoisin. Role of the second and third intracellular loops of metabotropic glutamate receptors in mediating dual signal transduction activation. *J Biol Chem* **1998**; *273*: 5615–5624.

30 Duthey, B., S. Caudron, J. Perroy, B. Bettler, L. Fagni, J.-P. Pin, L. Prézeau. A single subunit (GB2) is required for G-protein activation by the heterodimeric GABAB receptor. *J Biol Chem* **2002**; *277*: 3236–3241.

31 Gomeza, J., C. Joly, R. Kuhn, T. Knopfel, J. Bockaert, J.-P. Pin. The second intracellular loop of mGluR1 cooperates with the other intracellular domains to control coupling to G protein. *J Biol Chem* **1996**; *271*: 2199–2205.

32 De Blasi, A., P.J. Conn, J.-P. Pin, F. Nicoletti. Molecular determinants of metabotropic glutamate receptor signaling. *Trends Pharmacol Sci* **2001**; *22*: 114–120.

33 Couve, A., P. Thomas, A.R. Calver, W.D. Hirst, M.N. Pangalos, F.S. Walsh, T.G. Smart, S.J. Moss. Cyclic AMP-dependent protein kinase phosphorylation facilitates GABA(B) receptor-effector coupling. *Nat Neurosci* **2002**; *5*: 415–424.

34 Yamashita, T., A. Terakita, Y. Shichida. The second cytoplasmic loop of metabotropic glutamate receptor functions at the third loop position of rhodopsin. *J Biochem* **2001**; *130*: 149–155.

35 Pin, J.-P., C. Joly, S.F. Heinemann, J. Bockaert. Domains involved in the specificity of G-protein activation in phospholipase C coupled metabotropic glutamate receptors. *EMBO J* **1994**; *13–2*: 342–348.

36 Palczewski, K., et al. Crystal structure of rhodopsin: A G protein-coupled receptor (see comments). *Science* **2000**; *289*: 739–745.

37 Jensen, A.A., T.A. Spalding. Allosteric modulation of G-protein coupled receptors. *Eur J Pharm Sci* **2004**; *21*: 407–420.

38 Malherbe, P., et al. Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, non-competitive antagonist of the metabotro-

pic glutamate 1 receptor. *J Biol Chem* **2003**; *278*: 8340–8347.

39 Malherbe, P., N. Kratochwil, M.T. Zenner, J. Piussi, C. Diener, C. Kratzeisen, C. Fischer, R.H. Porter. Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. *Mol Pharmacol* **2003**; *64*: 823–832.

40 Pagano, A., et al. The non-competitive antagonists 2-methyl-6-(phenylethynyl)-pyridine and 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester interact with overlapping binding pockets in the transmembrane region of group I metabotropic glutamate receptors. *J Biol Chem* **2000**; *275*: 33750–33758.

41 Goudet, C., V. Binet, L. Prezeau and J.-P. Pin. Allosteric modulators of class C G protein coupled receptors open new possibilities for therapeutical applications. *Drug Discovery Today: Ther Strat* **2004**; *1*: 125–133.

42 Kew, J.N. Positive and negative allosteric modulation of metabotropic glutamate receptors: emerging therapeutic potential. *Pharmacol Ther* **2004**; *104*: 233–244.

43 Xu, H., L. Staszewski, H. Tang, E. Adler, M. Zoller, X. Li. Different functional roles of T1R subunits in the heteromeric taste receptors. *Proc Natl Acad Sci USA* **2004**; *101*: 14258–14263.

44 Binet, V., C. Brajon, L. Le Corre, F. Acher, J.P. Pin, L. Prézeau. The heptahelical domain of GABAB2 is directly activated by CGP7930, a positive allosteric modulator of the GABAB receptor. *J Biol Chem* **2004**; *279*: 29085–29091.

45 Fagni, L., F. Ango, J. Perroy, J. Bockaert. Identification and functional roles of metabotropic glutamate receptor-interacting proteins. *Semin Cell Dev Biol* **2004**; *15*: 289–298.

46 Romano, C., W.-L. Yang, K.L. O'Malley. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. *J Biol Chem* **1996**; *271*: 28612–28616.

47 Ray, K., B.C. Hauschild, P.J. Steinbach, P.K. Goldsmith, O. Hauache,

A.M. Spiegel. Identification of the cysteine residues in the amino-terminal extracellular domain of the human Ca^{2+} receptor critical for dimerization. Implications for function of monomeric Ca^{2+} receptor. *J Biol Chem* **1999**; *274*: 27642–27650.

48 Ray, K., B.C. Hauschild. Cys-140 Is critical for metabotropic glutamate receptor-1 (mGluR-1) dimerization. *J Biol Chem* **2000**; *275*: 34245–34251.

49 Tsuji, Y., et al. Cryptic dimer interface and domain organization of the extracellular region of metabotropic glutamate receptor subtype 1. *J Biol Chem* **2000**; *275*: 28144–28151.

50 Romano, C., J.K. Miller, K. Hyrc, S. Dikranian, S. Mennerick, Y. Takeuchi, M.P. Goldberg, K.L. O'Malley. Covalent and noncovalent interactions mediate metabotropic glutamate receptor mGlu5 dimerization. *Mol Pharmacol* **2001**; *59*: 46–53.

51 Jones, K.A., et al. GABA B receptors function as a heteromeric assembly of the subunits GABA B R1 and GABA B R2. *Nature* **1998**; *396*: 674–679.

52 White, J.H., et al. Heterodimerisation is required for the formation of a functional GABA B receptor. *Nature* **1998**; *396*: 679–682.

53 Kuner, R., G. Köhr, S. Grünewald, G. Eisenhardt, A. Bach, H.-C. Kornau. Role of heteromer formation in GABAB receptor function. *Science* **1999**; *283*: 74–77.

54 Kaupmann, K., et al. GABA B-receptor subtypes assemble into functional heteromeric complexes. *Nature* **1998**; *396*: 683–687.

55 Kniazeff, J., T. Galvez, G. Labesse, J.-P. Pin. No ligand binding in the GB2 subunit of the GABAB receptor is required for activation and allosteric interaction between the subunits. *J Neurosci* **2002**; *22*: 7352–7361.

56 Margeta-Mitrovic, M., Y.N. Jan, L.Y. Jan. Function of GB1 and GB2 subunits in G protein coupling of GABA(B) receptors. *Proc Natl Acad Sci USA* **2001**; *98*: 14649–14654.

57 Robbins, M.J., et al. GABAB2 is essential for G-protein coupling of the

GABAB receptor heterodimer. *J Neurosci* **2001**; *21*: 8043–8052.

58 Gassmann, M., et al. Redistribution of GABAB(1) protein and atypical GABAB responses in GABAB(2)-deficient mice. *J Neurosci* **2004**; *24*: 6086–6097.

59 Schuler, V., et al. Epilepsy, hyperalgesia, impaired memory, and loss of pre- and postsynaptic gaba(b) responses in mice lacking gaba(b(1)). *Neuron* **2001**; *31*: 47–58.

60 Prosser, H.M., et al. Epileptogenesis and enhanced prepulse inhibition in gaba(b1)-deficient mice. *Mol Cell Neurosci* **2001**; *17*: 1059–1070.

61 Thuault, S.J., et al. The GABA(B2) subunit is critical for the trafficking and function of native GABA(B) receptors. *Biochem Pharmacol* **2004**; *68*: 1655–1666.

62 Margeta-Mitrovic, M., Y.N. Jan, L.Y. Jan. A trafficking checkpoint controls GABAB receptor heterodimerization. *Neuron* **2000**; *27*: 97–106.

63 Pagano, A., et al. C-terminal interaction is essential for surface trafficking but not for heteromeric assembly of GABAB receptors. *J Neurosci* **2001**; *21*: 1189–1202.

64 Nelson, G., M.A. Hoon, J. Chandrashekar, Y. Zhang, N.J.P. Ryba, C.S. Zuker. Mammalian sweet taste receptors. *Cell* **2001**; *106*: 381–390.

65 Nelson, G., J. Chandrashekar, M.A. Hoon, L. Feng, G. Zhao, N.J. Ryba, C.S. Zuker. An amino-acid taste receptor. *Nature* **2002**; *416*: 199–202.

66 Kniazeff, J., P.-P. Saintot, C. Goudet, J. Liu, A. Charnet, G. Guillon, J.-P. Pin. Locking the dimeric GABAB G-protein coupled receptor in its active state. *J Neurosci* **2004**; *24*: 370–377.

67 Kubo, Y., T. Miyashita, Y. Murata. Structural basis for a Ca^{2+}-sensing function of the metabotropic glutamate receptors. *Science* **1998**; *279*: 1722–1725.

68 Sato, T., Y. Shimada, N. Nagasawa, S. Nakanishi, H. Jingami. Amino acid mutagenesis of the ligand binding interface of the metabotropic glutamate receptor crucial residues for setting the activated state. *J Biol Chem* **2003**; *278*: 4314–4321.

69 LIU, J.F., D. MAUREL, S. ETZOL, I. BRABET, H. ANSANAY, J.P. PIN, P. RONDARD. Molecular determinants of the allosteric control of agonist affinity in GABAB receptor by the GABAB2 subunit. *J Biol Chem* **2004**; *279*: 15824–15830.

70 GALVEZ, T., B. DUTHEY, J. KNIAZEFF, J. BLAHOS, G. ROVELLI, B. BETTLER, L. PRÉZEAU, J.-P. PIN. Allosteric interactions between GB1 and GB2 subunits are required for optimal GABAB receptor function. *EMBO J* **2001**; *20*: 2152–2159.

71 MARGETA-MITROVIC, M., Y.N. JAN, L.Y. JAN. Ligand-induced signal transduction within heterodimeric GABA(B) receptor. *Proc Natl Acad Sci USA* **2001**; *98*: 14643–14648.

72 GALVEZ, T., et al. Mapping the agonist binding site of GABAB type 1 subunit sheds light on the activation process of GABAB receptors. *J Biol Chem* **2000**; *275*: 41166–41174.

73 LI, X., L. STASZEWSKI, H. XU, K. DURICK, M. ZOLLER, E. ADLER. Human receptors for sweet and umami taste. *Proc Natl Acad Sci USA* **2002**; *99*: 4692–4696.

74 KNIAZEFF, J., A.-S. BESSIS, D. MAUREL, H. ANSANAY, L. PREZEAU, J.-P. PIN. Closed state of both binding domains of homodimeric mGlu receptors is required for full activity. *Nat Struct Mol Biol* **2004**; *11*: 706–713.

75 CARROLL, F.Y., et al. BAY36–7620: a potent non-competitive mGlu1 receptor antagonist with inverse agonist activity. *Mol Pharmacol* **2001**; *59*: 965–973.

76 MATHIESEN, J.M., N. SVENDSEN, H. BRAUNER-OSBORNE, C. THOMSEN, M.T. RAMIREZ. Positive allosteric modulation of the human metabotropic glutamate receptor 4 (hmGluR4) by SIB-1893 and MPEP. *Br J Pharmacol* **2003**; *138*: 1026–1030.

77 O'BRIEN, J.A., et al. A family of highly selective allosteric modulators of the metabotropic glutamate receptor subtype 5. *Mol Pharmacol* **2003**; *64*: 731–740.

78 O'BRIEN, J.A., et al. A novel selective allosteric modulator potentiates the activity of native metabotropic glutamate receptor subtype 5 in rat forebrain. *J Pharmacol Exp Ther* **2004**; *309*: 568–577.

79 PINKERTON, A.B., et al. Allosteric potentiators of the metabotropic glutamate receptor 2 (mGlu2). Part 3: Identification and biological activity of indanone containing mGlu2 receptor potentiators. *Bioorg Med Chem Lett* **2005**; *15*: 1565–1571.

80 JOHNSON, M.P., et al. Discovery of allosteric potentiators for the metabotropic glutamate 2 receptor: synthesis and subtype selectivity of N-(4-(2-methoxyphenoxy)phenyl)-N-(2,2,2- trifluoroethylsulfonyl)pyrid-3-ylmethylamine. *J Med Chem* **2003**; *46*: 3189–3192.

81 SCHAFFHAUSER, H., et al. Pharmacological characterization and identification of amino acids involved in the positive modulation of metabotropic glutamate receptor subtype 2. *Mol Pharmacol* **2003**; *64*: 798–810.

82 KNOFLACH, F., V. MUTEL, S. JOLIDON, J.N. KEW, P. MALHERBE, E. VIEIRA, J. WICHMANN, J.A. KEMP. Positive allosteric modulators of metabotropic glutamate 1 receptor: Characterization, mechanism of action, and binding site. *Proc Natl Acad Sci USA* **2001**; *98*: 13402–13407.

83 GOUDET, C., et al. Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. *Proc Natl Acad Sci USA* **2004**; *101*: 378–383.

84 URWYLER, S., T. GJONI, J. KOLJATIC, D.S. DUPUIS. Mechanisms of allosteric modulation at GABA(B) receptors by CGP7930 and GS39783: effects on affinities and efficacies of orthosteric ligands with distinct intrinsic properties. *Neuropharmacology* **2005**; *48*: 343–353.

85 PETREL, C., A. KESSLER, F. MASLAH, P. DAUBAN, R.H. DODD, D. ROGNAN, M. RUAT. Modeling and mutagenesis of the binding site of Calhex 231, a novel negative allosteric modulator of the extracellular Ca(2+)-sensing receptor. *J Biol Chem* **2003**; *278*: 49487–49494.

86 MIEDLICH, S.U., L. GAMA, K. SEUWEN, R.M. WOLF, G.E. BREITWIESER. Homology modeling of the transmembrane domain of the human calcium sensing receptor and localization of an allosteric

binding site. *J Biol Chem* **2004**; *279*: 7254–7263.

87 Ray, K., J. Northup. Evidence for distinct cation and calcimimetic compound (NPS 568) recognition domains in the transmembrane regions of the human Ca^{2+} receptor. *J Biol Chem* **2002**; *277*: 18908–18913.

88 Joubert, L., S. Claeysen, M. Sebben, A.-S. Bessis, R.D. Clark, R.S. Martin, J. Bockaert, A. Dumuis. A 5-HT4 receptor transmembrane network implicated in the activity of the inverse agonists but not agonists. *J Biol Chem* **2002**; *277*: 25502–25511.

89 Okada, T., O. Enrnst, K. Palczewski, K.P. Hofmann. Activation of rhodopsin: new insights from structural and biochemical studies. *Trends Biochem Sci* **2001**; *26*: 318–324.

90 Tateyama, M., H. Abe, H. Nakata, O. Saito, Y. Kubo. Ligand-induced rearrangement of the dimeric metabotropic glutamate receptor 1alpha. *Nat Struct Mol Biol* **2004**; *11*: 637–642.

91 Park, P.S., S. Filipek, J.W. Wells, K. Palczewski. Oligomerization of G protein-coupled receptors: past, present, and future. *Biochemistry* **2004**; *43*: 15643–15656.

92 Hamm, H.E. How activated receptors couple to G proteins. *Proc Natl Acad Sci USA* **2001**; *98*: 4819–4821.

93 Liang, Y., D. Fotiadis, S. Filipek, D.A. Saperstein, K. Palczewski, A. Engel. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem* **2003**; *278*: 21655–21662.

94 Baneres, J.L., A. Martin, P. Hullot, J.P. Girard, J.C. Rossi, J. Parello. Structure-based analysis of GPCR function: conformational adaptation of both agonist and receptor upon leukotriene B(4) binding to recombinant BLT1. *J Mol Biol* **2003**; *329*: 801–814.

95 Baneres, J.-L., J. Parello. Structure-based analysis of GPCR function. Evidence for a novel pentameric assembly between the dimeric leukotriene B(4) receptor BLT1 and the G-protein. *J Mol Biol* **2003**; *329*: 815–829.

96 Kuhn, H. Early steps in the light-triggered activation of the cyclic GMP enzymatic pathway in rod photoreceptors. *Prog Clin Biol Res* **1984**; *164*: 303–311.

97 Havlickova, M., J. Blahos, I. Brabet, J. Liu, B. Hruskova, L. Prezeau, J.P. Pin. The second intracellular loop of metabotropic glutamate receptors recognizes C termini of G-protein alpha-subunits. *J Biol Chem* **2003**; *278*: 35063–35070.

98 Hlavackova, V., et al. Evidence for a single heptahelical domain being turned on upon activation of a dimeric GPCR. *EMBO J* **2005**; *24*, 499–509.

99 El-Asmar, L., J.Y. Springael, S. Ballet, E.U. Andrieu, G. Vassart, M. Parmentier. Evidence for negative binding cooperativity within CCR5-CCR2b heterodimers. *Mol Pharmacol* **2005**; *67*: 460–469.

5
Molecular Mechanisms of GPCR Activation

Robert P. Bywater and Paul Denny-Gouldson

5.1
Structure of G Protein-coupled Receptors

G protein-coupled receptors (GPCRs) are the major signal transduction proteins in higher eukaryotes. Like any other signaling system, GPCRs are expected to have the facility to switch between two (or more) states in order to transmit the signal. This "two-state" mechanism is not unique to GPCRs since many proteins switch between active and inactive states. The switch is often through rigid body movements of one part of the structure relative to another about a molecular hinge [1]. This has been shown to be true in the case of enzymes, and a two-state enzyme will be considered in this chapter, the G protein, whose switching is concerted by the corresponding switch in the GPCR itself. So, for both GPCRs and for G proteins, two conformational states need to be considered.

Before proceeding to a discussion of the mechanisms of receptor action, a few brief words about structure are in order (for a more detailed discussion the reader is referred to earlier chapters in this book). The heptahelical topology of GPCRs is now well established and has been experimentally confirmed by the crystal structure determination of bovine rhodopsin independently by two groups [2–6]. While most of the work reviewed here has used the 1f88 structure [4] as the template for modeling and design work, we consider the 1gzm structure [5] to be a better choice, on account of the more detailed structure of the loop regions and the higher resolution throughout.

This chapter focuses on aspects of GPCR structure important for ligand binding, activation and G-protein binding. An important point to be made at the outset is that the crystal structures obtained so far [2–4] correspond to the dark (inactive) state of rhodopsin. The assumption that will be made here is that there are two main conformational states accessible to GPCRs [7], the active and the inactive states, often referred to as the R* and R states, respectively [8]. As a useful step towards the determination of the structure of the active state of a GPCR, the structure of metarhodopsin I has been determined [6]. This structure is distinct from the ground state, although similar to it. It is however not exactly the state that

would correspond to R*, a closer analog would be the metarhodopsin II state. This has so far eluded crystallization. A further point to remember is that, while the crystal structure of rhodopsin has immeasurably improved our understanding of GPCR structure in general, there are reasons for exercising caution in using the coordinates of rhodopsin as a template completely uncritically [9].

This chapter will be confined to a consideration of the Class A receptors, those related by sequence to bovine rhodopsin. While it is commonly assumed that Class B and C GPCRs possess a transmembrane fold essentially the same as the Class A receptors, to the extent that ligands can bind in the same way to different GPCR classes [10], the Class B and C receptors deserve to be dealt with in separate reviews.

5.2
Activation of GPCRs by Endogenous Ligands: The Concept of Receptor Agonism

The identity of the residues important for recognition of the endogenous ligands for GPCRs were elucidated by many workers during the early 1990s and has been the subject of many reviews [11] as well as other chapters in this book. This ligand binding region encompasses inward-facing residues on TMs 3, 5, 6 and 7.

Further, the classical site has been extended to include TM4, now shown to be important for ligand binding in some receptor subtypes such as cannabinoid [12] and histamine [13], and to be accessible to water-soluble probes [14].

A second binding pocket, formed by TMs 1, 2, 3 and 7 was considered to be the binding site for antagonists [15], but here again, the picture is very complicated. Both peptides and other large ligands have been shown to utilize this secondary site, and external loop structures, as well as the classical site. Furthermore, these large peptide ligands are agonists, so the concept that small ligands are agonists and larger ones are antagonists does not hold for all GPCRs. And, yet further, the very concept of "antagonist" needs to be defined more clearly (cf. further discussion below).

5.3
Distinction Between Orthosteric and Allosteric Ligands

There has been much discussion in the pharmacology literature [16–20] concerning allosteric, as opposed to orthosteric ligands. It would be convenient to be able to define a common orthosteric site, but this seems difficult in the light of the complex manner in which different ligand classes bind. Probably, the best definition is the pharmacological one – an allosteric ligand does not compete with the endogenous agonist, while an orthosteric ligand does.

This distinction raises one critical issue concerning the use of purely ligand-based approaches in the design of new ligands. Frequently, the set of ligands used to design the pharmacophore bind in different sites although all show the same

pharmacological effect, say antagonism. But because there is no information about the target built into these models, serious errors can be introduced. It is therefore essential that modelers involved in drug design be very critical and detailed in their analysis of the biological data used.

5.4
Only a Few Receptor Types are Known to Possess an Endogenous Antagonist

The members of the melanocortin receptor family are distributed both peripherally e.g. MC1R in skin, with an important role in skin pigmentation, and, in the case of MC4R, in the CNS, where it has a pivotal role in control of feeding behavior and energy balance [21]. The endogenous agonist is α-melanocyte-stimulating hormone [22], but the melanocortin receptors also possess an endogenous competitive antagonist, the Agouti-related peptide [23, 24].

To date, only one other family of GPCRs has been shown to be subject to an antagonist effect by a natural ligand. These are the "pH sensitive" GPCRs which are pH sensitive in a functionally very important way, not simply by reacting to a pH change. For example, the main function of TDAG8 is not just responding to proinflammatory lipids, but, rather, regulating cellular responses in acidic microenvironments [25]. TDAG8 may play a biological role in immune response and cellular transformation under conditions accompanying tissue acidosis. These proton-sensing G protein-coupled receptors include TDAG8, GPR4, OGR1, and GPR4 [26, 27], and many of them are oncogenic and overexpressed in human cancers [28]. These regulatory responses are clearly very important. The (lyso)phospholipid mediators sphingosine-1-phosphate, lysophosphatidic acid, sphingosylphosphorylcholine, and phosphatidic acid have been reported [29] to regulate *inter alia* diverse cellular responses such as proliferation, survival and death, cytoskeletal rearrangements, cell motility, and differentiation. Other studies [30] show that sphingosine-1-phosphate and lysophosphatidic acid are not the ligands for GPR4. Some of the endogenous ligands for this receptor behave antagonistically, as exemplified by the observation [31] that psychosine and related lysosphingolipids behave as if they were antagonists against proton-sensing receptors, including TDAG8 and the related receptor types GPR4, and OGR1. Similarly, it was shown [32] that G2A is a proton-sensing G protein-coupled receptor antagonized by lysophosphatidylcholine.

The concept of receptor antagonism is not new, but these cases illustrate that antagonists can be truly competitive for endogenous agonist. Such antagonists are better referred to as inverse agonists [33–38].

Other classes of antagonist can act by binding at other sites, for which a better term would be allosteric inhibitor (or attenuator).

5.5
Constitutively Active GPCRs

Most GPCRs exhibit some activity in the absence of ligand i. e. at zero drug concentration a residual receptor activity is observed. The receptor is said to be constitutively active. The level of this basal activity varies from receptor type to receptor type. This is a naturally-occurring event that confers advantages to the organism. But it may in the cases of certain mutations be an undesirable event leading to a disease state [39, 40].

Many mechanisms have been suggested for constitutive activity caused by mutations (CAM) but the general consensus is that the mutation(s) in a CAM destabilize(s) the inactive state of the receptor and/or stabilize(s) the active state of the receptor. This has the effect of increasing the number of receptor molecules capable of attaining the active conformation, giving rise to the basal level in activity assays as described above. A number of studies have shown that these mutations are not restricted to one region of the receptor but can occur throughout the structure.

The location of the mutated residues provides valuable information that aids in the understanding of the mechanism of activation of GPCRs. There is strong evidence that TM3 and TM6 move apart on activation [41–49], and it comes as no surprise that many cases of constitutive activity arise as a result of mutations in the lower regions of these TMs, e. g. residues 339 [41], 610 [50], 620 and 625 [51], in the DRY motif [44], in TM6 [52] and in TMS 3 and 7 [53].

Other factors that can affect basal activity include levels of steroids, sugars, and lipid composition [54, 55], and overexpression can have this effect also. This is often overlooked when comparing activity results for a receptor expressed in different cell types. It should also be noted that the effect of these environmental variables is not the same for all GPCRs; some are "deactivated" where others are "activated".

5.6
Mechanism of GPCR Activation: The Active/Inactive "Switch"

Until recently it was only one state of the receptor which was considered when modeling GPCRs [7]. All models produced were used in docking experiments where often both agonists and antagonists were docked into one structure. An illuminating example of this is the study [56] which concluded that the rhodopsin structure is suitable as a template for designing antagonists, but of only limited value for designing agonists. It should be recalled in this context that the rhodopsin crystal structure is the inactive form of that receptor.

Modeling two states of the receptor allows hypotheses to be generated and tested as to the structural changes that occur upon activation. There is an abundance of published biochemical data available on active receptor conformations and using this has enabled the changes that occur to be modeled [7]. The biochemical data from spin-labeling studies [45, 46, 49] suggests that there is move-

ment of TM6 away from TM3 at the intracellular side of the receptor, with some changes also in the tilt of TM3. Substituted cysteine accessibility data [57] suggest that TM6 either rotates, or there is a rearrangement within the TM bundle, to expose Cys(619) to the aqueous environment upon activation. Disulfide cross-linking data [58, 59] suggest that the intracellular end of TM7 and TM3 undergo rearrangement, along with TM5 and TM3. Further evidence that TM3 and TM6 move apart on activation [41–44, 47, 48] has been alluded to previously.

It is also possible to derive restraints from ligands, and this has been done for the agonist epinephrine in the beta-2-adrenergic receptor [7]. This ligand-based restraint information suggests that the distance between TM3 and TM5 at the extracellular side of the receptor decreases during activation relative to that seen in 1gzm-based inactive models.

Some groups have suggested that the activation process involves changes of conformation around prolines and conserved sequence motifs [60, 61]. Another mechanism suggested for activation is side-chain rotamer shifts. A good example is that of the R of DRY, which in 1gzm is in a sparsely populated rotamer. There is further experimental evidence that supports the rotamer theory, however they are by no means the sole method of activation, as large structural changes affecting the backbone geometry also occur [62].

It is interesting to consider which residue positions are responsible for these changes in backbone structure. It is expected that the "switch" residues will be very highly conserved, while the ligand-specific residues will be only class-conserved. Other residue positions, responsible for other functional duties such as in-
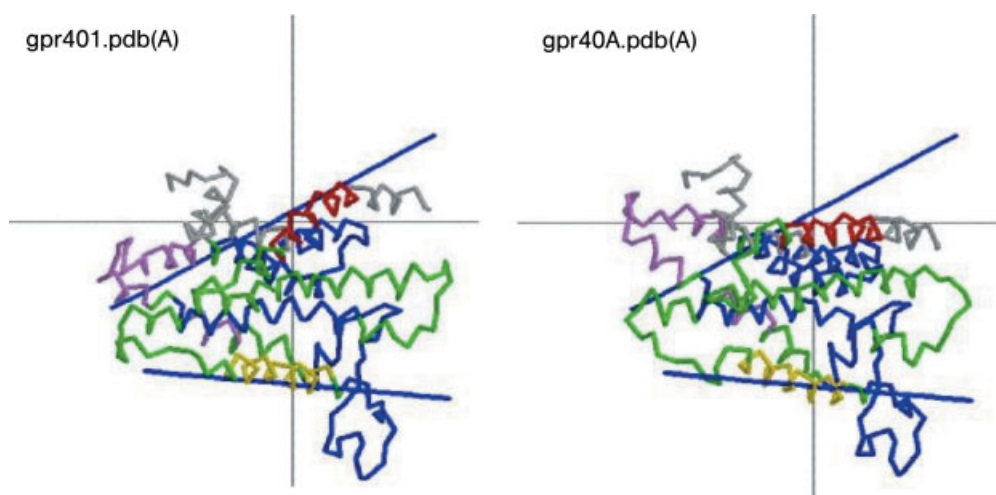

Fig. 5.1 Cartoons showing inactive (left) and active (right) structures of GPR40 built on the templates described in [7]. Note that the two structures are related by a rigid-body movement about a hinge region, the two axes of the rigid bodies are marked with a solid black line. Figures were obtained by using the program DynDom [1].

tegration into the membrane are probably less conserved [63, 64]. One way to map the level of conservation is to consider the entropy of residue-type appearance at a given locus [63, 64]. Using these methods, it is easy to demonstrate that conservation is distributed in such a way as to preserve function: general function for the "switch" mechanism, more specific (class-conserved) for ligands, and so on.

To summarize, changes are expected to be seen in all the TM domains, which may consist of side-chain rotamer changes, helix tilt, rotation and translation movements. Fig. 5.1 shows a comparison of an active and inactive model for a typical Class A GPCR to highlight the predicted hinge position, the axes of the rigid body elements that move in relation to the hinges and changes in relative helix positions.

The next step for GPCR scientists is to endeavor to understand how an agonist stabilizes the active conformation and conversely, what do antagonists and, importantly, inverse agonists do. Currently the GPCR modeling community is only just beginning to investigate this through the use of active and inactive models, and it appears that the system is very complicated.

One example of a complicating factor is partial agonism. It is not yet understood whether this is due to low receptor occupancy or inability of the ligand to stabilize the conformation associated with the full agonist. Other conformations, in addition to the R and R* states cannot be ruled out either. Several intermediate states are known to exist in rhodopsin [65, 66] and an intermediate (R′) state can be observed in a mutant beta-2-adrenergic receptor [67], and there is no *a priori* reason why the same should not apply to other GPCRs that are related in sequence and structure to rhodopsin.

An additional factor to be taken into consideration is that the changes associated with activation as part of the phosphorylation/downregulation pathway have not yet been elucidated. Some ligands are able to activate receptors without causing internalization and vice versa, so more work is required to delineate the different conformational signals.

5.7
GPCR Dimerization

The fact that GPCRs form dimers or higher oligomers is well established experimentally using biochemical/molecular pharmacology approaches [68–72] and structure-based methods [68, 73–75]. Indeed, electron microscope data showing the presence of dimers was published even earlier [76–78], even though the presence of dimers was mentioned explicitly in only one of these papers [78].

Many of these observations have been rationalized and the ideas further developed using bioinformatics and computer modeling [79–82].

It is not clear from any of the foregoing experiments or calculations, which site(s) on the surface of the GPCRs form the dimer interface. The 5:6 dimer interface has figured prominently in earlier work [70, 80–82]. Other interfaces have been proposed [80] and observed e.g. the interfaces between TMs 4 and 5, the

third intracellular loop and TMs 1 and 2, and the interface between TM1 and its counterpart on a neighboring molecule, detected by atomic force microscopy [74]. Multiple sites of interaction have been invoked elsewhere [83]. What is striking about many of the interactions observed by structural methods (crystallography, atomic force microscopy) is that there does not seem to be a large interface between the participating molecules typical of functional dimers. This suggests that they may not be functional interfaces but crystal packing interactions, not important for function.

From the point of view of function, it is important to know whether the dimeric species is a necessary agent on the signal transduction pathway. Mathematical models of GPCR kinetics [84] strongly support this notion. Experimental confirmation has been provided by ligand-binding studies under conditions of varying receptor concentration [68] and by isothermal titration calorimetry and sedimentation equilibrium techniques [85]. It has been shown by both chromatographic and ultracentrifugation methods that both the inactive (dark) and active (light) forms of bovine rhodopsin are dimeric [73]. The dimer concept thus seems well established and this means that there will be profound effects on the pharmacology of GPCR action [86]. Apart from anything else, an agonist at sufficiently high concentration will, according to the dimer model, behave as an antagonist [84].

In addition to homodimers, heterodimers have been observed [87, 88]. The occurrence of heterodimers will of course vastly increase the functional repertoire of the GPCRs [89, 90]. Domain-swapped dimer models [82, 91] have been proposed to explain the observations of cross-talk and functional rescue in twinned GPCR chimeras [72]. It is hard to think of any other convincing explanation for these phenomena, and domain swapping has the advantage that well-packed internal interfaces that have already been evolutionarily created can be re-used in formation of the domain-swapped dimer.

Further, it is not certain that only dimers are involved, higher oligomers may also be involved [75, 92, 93]. It is not entirely clear what the functional advantages of oligomerization beyond the dimer could be. Whatever the case may be, it is in keeping with the principle of Ockham's razor to establish the simplest model that explains the observed facts.

5.8
Activation of G Proteins

While there is only a very restricted set of crystal structures of GPCRs (essentially rhodopsin in ground state and meta I, but not in a truly active (R*) state), there is a number of crystal structures of G proteins in both active and inactive states. An example of such a pair of structures, from the alpha subunits of the heterotrimeric G proteins of the G_i family, is shown in Fig. 5.2. Even a cursory inspection of these two structures and the differences between them shows that there are major conformational changes, most noticeably at the N-terminal chain, which undergoes
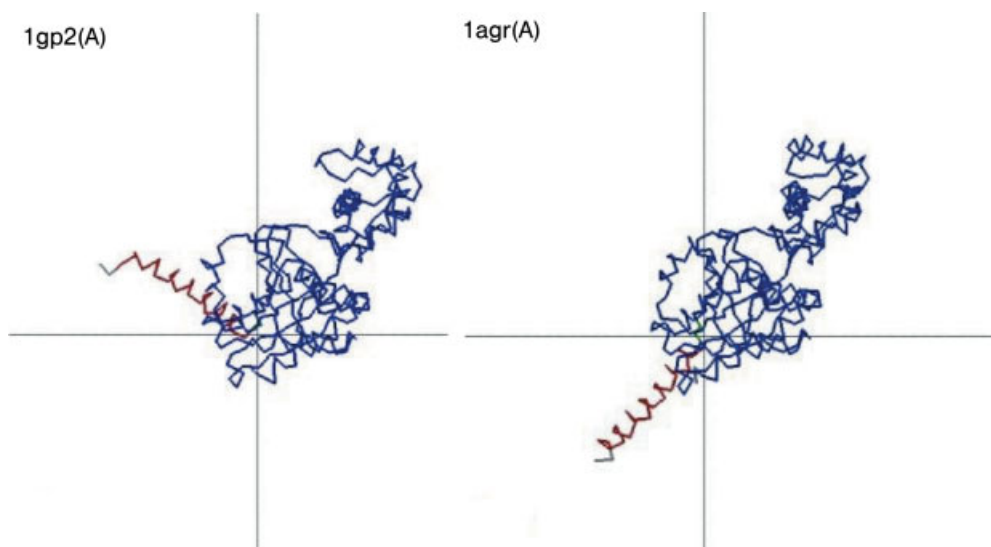
Fig. 5.2 Cartoons showing inactive (left) and active (right) structures of the rat G_i proteins G_i_α (PDB i.d. 1gp2a) and AlF4-Activated G_i-Alpha (PDB i.d. 1agra). Figures were obtained from the user database on the DynDom site [1].

an abrupt change in angle relative to the main body of the protein. The C-terminal regions also undergo significant changes, and this complies with published observations [94, 95]. These two regions of the G protein are more concerned with interactions with the beta-subunit of the heterotrimer. The same is true for the ligand (GDP) binding region; there are only very small conformational changes and displacements of the residues in this region. What this indicates is that, whatever small changes that do occur in the GPCR and ligand-binding regions somehow become amplified at the beta-subunit interface. This is consistent with the known fact that the beta-subunit is released as a result of G-protein activation by the GPCR. Only small differences between these two structures are discernible for the region where it is expected that the GPCR molecule(s) docks.

5.9
Interaction Between GPCRs and G Proteins

Proteins convey information by interacting with other proteins, and for the GPCRs, the principal partner is, by definition, a G protein. Normally, a given GPCR will interact with a G protein of only one of the main classes, G_s, G_i, G_o, and each of these has, in turn, its own set of partners in the subsequent downstream processing pathway.

As far as mechanism of action is concerned, one needs to know how the GPCR docks onto its designated G protein. A number of studies based on correlated mu-

tations [96] and evolutionary trace (ET) [80, 97, 98] have indicated important sites on the G proteins. The latter authors [97, 98] clearly demonstrate that the exposed "flat" region on the alpha subunit of the G-protein hetrotrimer is most likely to be involved in GPCR binding. The question of whether the GPCR docks as a monomer or dimer was not discussed by these authors, but in a later study [80] it was shown that the extent of the ET region on the "flat" surface was approximately twice as large as that previously detected, consistent with the notion that it is the dimer species that is the docking partner for the G protein.

As a further aid towards working out the details of which GPCR binds to which G-protein type, bioinformatics studies of GPCR–G-protein recognition have been performed [99] and an informative database of interactions between GPCRs and G proteins [100] is available on the internet.

Given that it is beginning to be possible to model GPCRs in both active and inactive conformations [7], and there are G protein crystal structures in both active and inactive states, at least for G_i, together with data which shows which residues on each partner, GPCR and G protein, are important for their mutual docking, it should be an easy matter to model this docking. There are two principle hindrances to actually achieving this in a convincing way. The first and most obvious one is that neither the models nor even the crystal structures of GPCRs are accurate enough in the region of the cytosolic loops. The second, more problematic question is the issue of stoichiometry. Does the GPCR dock onto the G protein as a dimer, as suggested by a substantial corpus of both theoretical and experimental data? These questions are open for future discussion, and above all, experimentation.

5.10
Conclusions

There is ample evidence that GPCRs can exist in at least two conformational states, active (R*) and inactive (R). This is not a unique feature of this class of proteins. Most, if not all, proteins exhibit active/inactive transitions. Yet it is surprising that many GPCR researchers seem to have ignored this. This can have serious consequences, because it can have profound consequences in drug design.

The primary targets for GPCR signal transduction are the G proteins. These also can exist in active or inactive states respectively. It is not known exactly how "active GPCR" binds to "active G protein" or whether "inactive GPCR" binds to "inactive G protein". This is an important issue to be settled. A further issue to be settled is stoichiometry. There is much evidence that the active form of GPCRs is dimeric (stoichiometry, ligand : receptor : G protein = 1 : 2 : 1). In that case, it is necessary to propose a mechanism by which a dimeric and therefore presumably symmetric complex can dock onto the G protein (heterotrimer) which must by its very nature be asymmetric. The issue of dimerization affects *inter alia* ligand kinetics. An agonist can, at high concentrations, behave as an antagonist. A greater awareness of such potential issues is necessary.

The residue positions in GPCRs that are involved in ligand recognition are becoming known. The residue positions that are involved in recognition of allosteric agents are also now beginning to be identified. However, the detailed mechanism of this switching process remains to be elucidated.

References

1 Hayward S, Berendsen HJC. Systematic analysis of domain motions in proteins from conformational change. *Proteins, Struct, Funct Genet* **1998**; *30*: 144–154. See also http://www.cmp.uea.ac.uk/dyndom

2 Teller DC, Okada T, Behnke CA, Palczewski K, Stenkamp RE. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors (GPCRs). *Biochemistry* **2001**; *40*: 7761–7772.

3 Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**; *289*. 739–745.

4 Palczewski K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

5 Li J, Edwards PC, Burghammer M, Villa, C, Schertler, G.F. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **2004**; *343*: 1409–1438.

6 Ruprecht JJ, Mielke T, Vogel R, Villa C, Schertler GFX. Electron crystallography reveals the structure of metarhodopsin I. *EMBO J* **2004**; *23*: 3609–3620.

7 Gouldson PR, Kidley NJ, Bywater RP, Psaroudakis G, Brooks HD, Diaz C, Shire D, Reynolds CA. Toward the active conformations of rhodopsin and the beta2-adrenergic receptor. *Proteins, Struct, Funct Genet* **2004**; *56*: 67–84.

8 Leff P. The two-state model of receptor activation. *TIPS* **1995**; *16*: 89–97.

9 Bywater RP. Location and nature of the residues important for ligand recognition in G-protein coupled receptors. *J Mol Recognit* **2005**; *18*: 60–72.

10 Sheikh SP, Vilardarga JP, Baranski TJ, Lichtarge, O, Iiri T, Meng EC, Nissenson RA, Bourne HR. Similar structures and shared switch mechanisms of the beta-2-adrenoceptor and the parathyroid hormone receptor. Zn(II) bridges between helices III and VI block activation. *J Biol Chem* **1999**; *274*: 17033–17041.

11 Strader CD, Fong TM, Tota MR, Underwood D, Dixon RAF. Structure and function of G protein-coupled receptors. *Ann Rev Biochem* **1994**; *63*: 101–132.

12 Gouldson P, Calandra B, Legoux P, Kerneis A, Rinaldi-Carmona M, Barth F, Le Fur G, Ferrara P, Shire D. Mutational analysis and molecular modelling of the antagonist SR 144528 binding site on the human cannabinoid CB(2) receptor. *Eur J Pharmacol* **2000**; *401*: 17–25.

13 Wieland K, Laak AM, Smit MJ, Kuhne R, Timmerman H, Leurs R. Mutational analysis of the antagonist-

binding site of the histamine H(1) receptor. *J Biol Chem* **1999**; *274*, 29994–30000.

14 Javitch JA, Shi L, Simpson MM, Chen J, Chiappa V, Visiers I, Weinstein H, Ballesteros JA. The fourth transmembrane segment of the dopamine D2 receptor: accessibility in the binding-site crevice and position in the transmembrane bundle. *Biochemistry* **2000**; *39*: 12190–12199.

15 Suryanarayana S, von Zastrow M, Kobilka BK. Identification of intramolecular interactions in adrenergic receptors. *J Biol Chem* **1992**; *267*: 21991–21994.

16 Soudijn W, Van Wijngaarden I, Ijzerman AP. Allosteric modulators of G-protein coupled receptors: perspectives and recent developments. *Drug Discovery Today* **2004**; *9*: 752–758.

17 May LT, Avlani VA, Sexton PM, Christopoulos A. Allosteric modulation of G protein-coupled receptors. *Curr Pharm Des* **2004**; *10*: 2003–2013.

18 Christopoulos A, Kenakin T. G-protein coupled receptors: allosterism and complexing. *Pharm Rev* **2002**; *54*: 323–374.

19 Birdsall NJ, Lazareno S, Popham A, Saldanha J. Multiple allosteric sites on muscarinic receptors. *Life Sci* **2001**; *68*: 2517–2524.

20 Järv J. A model of non-exclusive binding of agonist and antagonist on G-protein coupled receptors. *J Theoret Biol* **1995**; *175*: 577–582.

21 Seeley RJ, Drazen DL, Clegg DJ. The critical role of the melanocortin system in the control of energy balance. *Annu Rev Nutr* **2004**; *24*. 133–149.

22 Haskell-Luevano C, Nikiforovich G, Sharma SD, Yang YK, Dickinson C, Hruby VJ, Gantz I. Biological and conformational examination of stereochemical modifications using the template melanotropin peptide, Ac-Nle-c[Asp-His-Phe-Arg-Trp-Ala-Lys]-NH2, on human melanocortin receptors. *J Med Chem* **1997**; *40*: 1738–1748.

23. Pritchard LE, Armstrong D, Davies N, Oliver RL, Schmitz CA, Brennand JC, Wilkinson GF, White A. Agouti-related protein (83–132) is a competitive antagonist at the human melanocortin-4 receptor. *J Endocrinol* **2004**; *180*: 183–191.

24 Haskell-Luevano C, Lim S, Yuan W, Cone RD, Hruby VJ. Structure activity studies of the melanocortin antagonist SHU9119 modified at the 6, 7, 8, and 9 positions. *Peptides* **2000**; *21*: 49–57.

25 Radu CG, Nijagal A, McLaughlin J, Wang L, Witte ON. Differential proton sensitivity of related G protein-coupled receptors T cell death-associated gene 8 and G2A expressed in immune cells. *Proc Natl Acad Sci USA* **2005**; *102*: 1632–1637.

26 Ludwig MG, Vanek M, Guerini D, Gasser JA, Jones CE, Junker U, Hofstetter H, Wolf RM, Seuwen K. Proton-sensing G-protein-coupled receptors. *Nature* **2003**; *425*: 93–98.

27 Ishii S, Kihara Y, Shimizu T. Identification of T Cell Death-associated Gene 8 (TDAG8) as a Novel Acid Sensing G-protein-coupled Receptor. *J Biol Chem* **2005**; *280*: 9083–9087.

28 Sin WC, Zhang Y, Zhong W, Adhikarakunnathu S, Powers S, Hoey T, An S, Yang J. G protein-coupled receptors GPR4 and TDAG8 are oncogenic and overexpressed in human cancers. *Oncogene* **2004**; *23*: 6299–6303.

29 Kostenis E. Novel clusters of receptors for sphingosine-1-phosphate, sphingosylphosphorylcholine, and (lyso)-phosphatidic acid: new receptors for "old" ligands. *J Cell Biochem* **2004**; *92*: 923–936.

30 Bektas M, Barak LS, Jolly PS, Liu H, Lynch KR, Lacana E, Suhr KB, Milstien S, Spiegel S. The G protein-coupled receptor GPR4 suppresses ERK activation in a ligand-independent manner. *Biochemistry* **2003**; *42*: 12181–12191.

31 Wang JQ, Kon J, Mogi C, Tobo M, Damirin A, Sato K, Komachi M, Malchinkhuu E, Murata N, Kimura T, Kuwabara A, Wakamatsu K, Koizumi H, Uede T, Tsujimoto G, Kurose H, Sato T, Harada A, Misawa N, Tomura H, Okajima F. TDAG8 is a proton-sensing and psychosine-sensitive G-protein-coupled receptor. *J Biol Chem* **2004**; *279*: 45626–45633.

32 Murakami N, Yokomizo T, Okuno T, Shimizu T. G2A is a proton-sensing G-protein-coupled receptor antagonized by lysophosphatidylcholine. *J Biol Chem* **2004**; *279*: 42484–42491.

33 Dupre, D.J, Rola-Pleszczynski, M, Stankova, J. Inverse agonism: more than reverting constitutively active receptor signaling. *Biochem Cell Biol* **2004**; *82*: 676–680.

34 Kenakin T. Efficacy as a vector: the relative prevalence and paucity of inverse agonism. *Mol Pharmacol* **2004**; *65*: 2–11.

35 Strange PG. Mechanisms of inverse agonism at G-protein-coupled receptors. *Trends Pharmacol Sci* **2002**; *23*: 89–95.

36 De Ligt, R.A, Kourounakis, A.P, IJzerman, A.P. Inverse agonism at G protein-coupled receptors: (patho)physiological relevance and implications for drug discovery. *Br J Pharmacol* **2000**; *130*: 1–12.

37 Milligan G, Bond RA. Inverse agonism and the regulation of receptor number. *Trends Pharmacol Sci* **1997**; *18*: 468–474.

38 Chidiac P, Hebert TE, Valiquette M, Dennis, M, Bouvier M. Inverse agonist activity of beta-adrenergic antagonists. *Mol Pharmacol* **1994**; *45*: 490–499.

39 Seifert R, Wenzel-Seifert K. Constitutive activity of G-protein-coupled receptors: cause of disease and common property of wild-type receptors. *Naunyn Schmiedebergs Arch Pharmacol* **2002**; *366*: 381–416.

40 Parnot C, Miserey-Lenkei, S, Bardin, S, Corvol, P, Clauser, E. Lessons from constitutively active mutants of G-protein coupled receptors. *Trends Endocrinol Metabol* **2002**; *13*: 336–343.

41 Kim JM, Altenbach C, Kono M, Oprian DD, Hubbell WL, Khorana HG. Structural origins of constitutive activation in rhodopsin: Role of the K296/E113 salt bridge. *Proc Natl Acad Sci USA* **2004**; *101*: 12508–12513.

42 Ghanouni P, Steenhuis JJ, Farrens DL, Kobilka BK. Agonist-induced conformational changes in the G-protein-coupling domain of the beta-2-adrenergic receptor. *Proc Natl Acad Sci USA* **2001**; *98*: 5997–6002.

43 Ballesteros JA, Jensen AD, Liapakis G, Rasmussen SG, Shi L, Gether U, Javitch JA. Activation of the beta-2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* **2001**; *276*: 29171–29177.

44 Rasmussen SGF, Jensen AD, Lapiakis G, Ghanouni P, Javitch JA, Gether U. Mutation of a highly conserved aspartic acid in the $\alpha2$ adrenergic receptor: constitutive activation, structural instability and conformational rearrangement of transmembrane segment 6. *Mol Pharmacol* **1999**; *56*: 175–184.

45 Altenbach C, Yang K, Farrens DL, Farahbakhsh ZT, Khorana HG, Hubbell WL. Structural features and light-dependent changes in the cytoplasmic interhelical e–f loop region of rhodopsin – a site-directed spin-labeling study. *Biochemistry* **1996**; *35*: 12470–12478.

46 Altenbach, C, Cai K, Khorana HG, Hubbell WL. Structural features and light-dependent changes in the sequence 306–322 extending from helix VII to the palmitoylation sites in rhodopsin: a site-directed spin-labeling study. *Biochemistry* **1999**; *38*: 7931–7937.

47 Gether and Kobilka, G protein-coupled receptors. II Mechanism of against activation. *J Biol Chem* **1998**; *273*: 17979–17982.

48 Spalding TA, Burstein ES, Henderson SC, Ducote KR, Brann MR. Identification of a ligand-dependent switch within a muscarinic receptor. *J Biol Chem* **1998**; *273*: 21563–21568.

49 Altenbach C, Klein-Seetharaman J, Hwa J, Khorana HG, Hubbell WL. Structural features and light-dependent changes in the sequence 59–75 connecting helices I and II in rhodopsin: a site-directed spin-labeling study. *Biochemistry* **1999**; *38*: 7945–7949.

50 Han M, Smith SO, Sakmar TP. Constitutive activation of opsin by mutation of methionine 257 on transmembrane helix 6. *Biochemistry* **1998**; *37*: 8253–8261.

51 Sommers CM, Martin NP, Akal-Strader A, Becker JM, Naider F,

Dumont M. A limited spectrum of mutants causes constitutive activation of the yeast alpha-factor receptor. *Biochemistry* **2000**; *39*: 6898–6909.

52 Chen S, Lin F, Xu M, Graham RM. Phe303 in TM6 of the alpha1Badrenergic receptor is a key residue coupling TM helical movements to G-protein activation. *Biochemistry* **2002**; *41*: 588–596.

53 Befort K, Zilliox C, Filliol D, Yue SY, Kieffer BL. Constitutive activation of the mu-opioid receptor by mutations in transmembrane domains III and VII *J Biol Chem* **1999**; *274*: 18574–18581.

54 Wissink S, Van der Burg B, Katzenellenbogen B.S, Van der Saag PT. Synergistic activation of the serotonin-1A receptor by nuclear factor-kappa B and estrogen. *Mol Endocrinol* **2001**; *15*: 543–552.

55 Lagane B, Gaibelet G, Meilhoc E, Masson J-M, Cézanne L, Lopez A. Role of sterols in modulating the human μ-opioid receptor function in *Saccharomyces cerevisiae. Biol Chem* **2000**; *43*: 33197–33200.

56 Bissantz C, Bernard P, Hibert M, Rognan D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins, Struct, Funct Genet* **2003**; *50*: 5–25.

57 Javitch JA, Fu DY, Liapakis G, Chen JY. Constitutive activation of the beta(2) Adrenergic receptor alters the orientation of its sixth membrane-spanning segment. *J Biol Chem* **1997**; *272*: 18546–18549.

58 Kono M, Yu H, Oprian DD. Disulfide bond exchange in rhodopsin. *Biochemistry* **1998**; *37*: 1302–1305.

59 Yu H, Kono M, Oprian DD. State-dependent disulfide cross-linking in rhodopsin. *Biochemistry* **1999**; *38*: 12028–12032.

60 Huang P, Visiers I, Weinstein H, Liu-Chen LY. The local environment at the cytoplasmic end of TM6 of the mu-opioid receptor differs from those of rhodopsin and monoamine receptors: introduction of an ionic lock between the cytoplasmic ends of helices 3 and 6 by a L6.30(275)E mutation inactivates

the mu-opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. *Biochemistry* **2002**; *41*: 11972–11980.

61 Sansom MS, Weinstein H. Hinges, swivels and switches: the role of prolines in signalling *via* transmembrane alpha-helices. *Trends Pharmacol Sci* **2000**; *21*: 445–451.

62 Prioleau C, Visiers I, Ebersole BJ, Weinstein H, Sealfon SC. Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT2C receptor. Identification of a novel "locked-on" phenotype and double revertant mutations. *J Biol Chem* **2002**; *277*: 36577–36584.

63 Bondensgaard K, Ankersen M, Thøgersen H, Hansen BS, Wulff BS, Bywater RP. Recognition of privileged structures by G-protein coupled receptors. *J Med Chem* **2004**; *47*: 888–899.

64 Oliveira L, Paiva ACM, Vriend G. Correlated mutation analyses on very large sequence families. *Chembiochem* **2002**; *3*: 1010–1017.

65 Ritter E, Zimmermann K, Heck M, Hofmann KP, Bartl FJ. Transition of rhodopsin into the active metarhodopsin II state opens a new light-induced pathway linked to Schiff base isomerization. *J Biol Chem* **2004**; *279*: 48102–48111.

66 Vogel R, Ludeke S, Radu I, Siebert F, Sheves M. Photoreactions of metarhodopsin III. *Biochemistry* **2004**; *43*: 10255–10264.

67 Liapakis G, Chan C, Papadokostaki M, Javitch JA. Synergistic contributions of the functional groups of epinephrine to its affinity and efficacy at the beta2 adrenergic receptor. *Mol Pharmacol* **2004**; *65*: 1181–1190.

68 Baneres JL, Parello J. Structure-based analysis of GPCR function: evidence for a novel pentameric assembly between the dimeric leukotriene B4 receptor BLT1 and the G-protein. *J Mol Biol* **2003**; *329*: 815–329.

69 Hebert T, Loisel TP, Adam L, Ethier N, St. Onge S, Bouvier M. Functional rescue of a constitutively desensitized beta-2-AR through receptor dimerization. *Biochem J* **1998**; *330*: 287–293.

70 HEBERT TE, MOFFETT S, MORELLO JP, LOISEL TP, BICHET DG, BARRET C, BOUVIER M. A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation. *J Biol Chem* **1996**; *271*: 16384–16392.

71 MAGGIO R, BARBIER P, FORNAI F, CORSINI GU. Functional role of the third cytoplasmic loop in muscarinic receptor dimerization. *J Biol Chem* **1996**; *271*: 31055–31060.

72 MAGGIO R, VOGEL Z, WESS J. Coexpression studies with mutant muscarinic/adrenergic receptors provide evidence for intermolecular "cross-talk" between G-protein-linked receptors. *Proc Natl Acad Sci USA* **1993**; *90*: 3103–3107.

73 MEDINA R, PERDOMO D, BUBIS J. The hydrodynamic properties of dark- and light-activated states of n-dodecyl beta-D-maltoside-solubilized bovine rhodopsin support the dimeric structure of both conformations. *J Biol Chem* **2004**; *279*: 39565–39573.

74 LIANG Y, FOTIADIES D, FILIPEK S, SAPERTSTEIN DA, PALCZEWSKI K, ENGEL, A. Organization of the G-Protein coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem* **2003**; *278*: 21655–21662.

75 FOTIADIS D, LIANG Y, FILIPEK S, SAPERSTEIN DA, ENGEL A, PALCZEWSKI K. Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* **2003**; *421*: 127–128.

76 UNGER VM, SCHERTLER GFX. Low-resolution structure of bovine rhodopsin determined by electron cryomicroscopy. *Biophys J* **1995**; *68*: 1776–1786.

77 SCHERTLER GFX, VILLA C., HENDERSON R. Projection structure of rhodopsin. *Nature* **1993**; *362*: 770–772.

78 CORLESS JM, MCCASLIN DR, SCOTT BL. Two-dimensional rhodopsin crystals from disk membranes of frog retinal rod outer segments, *Proc Natl Acad Sci USA* **1982**; *80*: 1116–1120.

79 FILIZOLA M, WEINSTEIN H. Structural models for dimerization of G-protein coupled receptors: the opioid receptor homodimers. *Biopolymers* **2002**; *66*: 317–325.

80 DEAN MK, HIGGS C, SMITH R.E, BYWATER RP, SNELL CR, SCOT PD, UPTON, GJ, HOWE, TJ, REYNOLDS, C.A. Dimerization of G-protein-coupled receptors. *J Med Chem* **2001**; *44*: 4595–4614.

81 GOULDSON PR, DEAN MK, SNELL CR, BYWATER RP, GKOUTOS G, REYNOLDS CA. Lipid-facing correlated mutations and dimerization in G-protein coupled receptors. *Protein Eng* **2001**; *14*: 759–767.

82 GOULDSON PR, SNELL CR, BYWATER RP, HIGGS C, REYNOLDS CA. Domain swapping in G-protein coupled receptor dimers. *Protein Eng* **1998**; *11*: 1181–1193.

83 LEE SP, O'DOWD BF, RAJARAM RD, NGUYEN T, GEORGE SR. D2 dopamine receptor homodimerization is mediated by multiple sites of interaction, including an intermolecular interaction involving transmembrane domain 4. *Biochemistry* **2003**; *42*: 11023–11031.

84 BYWATER RP, SØRENSEN A, RØGEN P, HJORTH PG. Construction of the simplest model to explain complex receptor activation kinetics. *J Theoret Biol* **2002**; *218*: 139–147.

85 FERNANDO H, CHIN C, ROSGEN J, RAJARATHNAM K. Dimer dissociation is essential for interleukin-8 (IL-8) binding to CXCR1 receptor. *J Biol Chem* **2004**; *279*: 36175–36178.

86 MILLIGAN G. G protein-coupled receptor dimerization: function and ligand pharmacology. *Mol Pharmacol* **2004**; *66*: 1–7.

87 PERCHERANCIER Y, BERCHICHE Y, SLIGHT I, VOLKMER-ENGERT R, TAMAMURA H, FUJII N, BOUVIER M, HEVEKER N. Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers. *J Biol Chem* **2005**; Jan 4: [Epub ahead of print].

88 MANDRIKA I, PETROVSKA R, WIKBERG J. Melanocortin receptors form constitutive homo- and heterodimers. *Biochem Biophys Res Commun* **2005**; *326*: 349–354.

89 RIOS CD, JORDAN BA, GOMES I, DEVI LA. G-protein-coupled receptor dimerization: modulation of receptor function. *Pharmacol Ther* **2001**; *92*: 71–87.

90 GOMES I, JORDAN BA, GUPTA A, RIOS C, TRAPAIDZE N, DEVI LA. G-protein

coupled receptor dimerization: implications in modulating receptor function. *J Mol Med* **2001**; *79*: 226–242.

91 Bakker RA, Dees G, Carrillo JJ, Booth RG, Lopez-Gimenez JF, Milligan G, Strange PG, Leurs R. Domain swapping in the human histamine H1 receptor. *J Pharmacol Exp Ther* **2004**; *311*: 131–138.

92 Javitch JA. The ants go marching two by two: oligomeric structure of G-protein-coupled receptors. *Mol Pharmacol* **2004**; *66*: 1077–1082.

93 Salahpour A, Bonin H, Bhalla S, Petaja-Repo U, Bouvier' M. Biochemical characterization of beta2-adrenergic receptor dimers and oligomers. *Biol Chem* **2003**; *384*: 117–123.

94 Heydorn A, Ward RJ, Jorgensen R, Rosenkilde MM, Frimurer TM, Milligan G, Kostenis E. Identification of a novel site within G protein alpha subunits important for specificity of receptor-G protein interaction. *Mol Pharmacol* **2004**; *66*: 250–259.

95 Kostenis E, Martini L, Ellis J, Waldhoer M, Heydorn A, Rosenkilde MM, Norregaard PK, Jorgensen R, Whistler JL, Milligan G. A highly conserved glycine within linker I and the extreme C-terminus of G protein {alpha} subunits interact coopera-tively in switching GPCR-to-effector specificity. *J Pharmacol Exp Ther* **2005**; *313*: 78–87.

96 Horn F, Van der Wenden EM Oliveira, L Ijzerman AP, Vriend G. Receptors coupling to G proteins: is there a signal behind the sequence? *Proteins, Struct, Funct Genet* **2000**; *41*: 448–459.

97 Onrust R, Herzmark P, Chi P, Garcia PD, Lichtarge O, Kingsley C, Bourne HR. Receptor and betagamma binding sites in the alpha subunit of the retinal G protein transducin. *Science* **1997**; *275*: 381–384.

98 Lichtarge O, Bourne HR, Cohen FE. Evolutionarily conserved G alpha beta gamma binding surfaces support a model of the G protein–receptor complex. *Proc Natl Acad Sci USA* **1996**; *93*: 7507–7511.

99 Møller S, Vilo J, Croning MD. Prediction of the coupling specificity of G protein coupled receptors to their G proteins. *Bioinformatics* **2001**; *17*(Suppl 1): S174–S181.

100 Elefsinioti AL, Bagos PG, Spyropoulos IC, Hamodrakas SJ. A database for G proteins and their inter-action with GPCRs. *BMC Bioinformatics* **2004**; *5*: 208.

6
Allosteric Properties and Regulation of G Protein-coupled Receptors

Jean-Luc Galzi, Emeline Maillet, Sandra Lecat, Muriel Hachet-Haas,
Jacques Haiech, Marcel Hibert, and Brigitte Ilien

6.1
Introduction

In their original description by Monod Wyman and Changeux in 1965, allosteric proteins interconvert between discrete conformations endowed with distinct structural, pharmacological and functional properties [1]. This model, which is based on the resolution of hemoglobin structure [2], states that regulatory proteins exist as oligomers and spontaneously adopt at least two conformations, thereby accounting for the regulation of the biological activity of the protein and of its pharmacological properties. In addition, biological activity of regulatory proteins can be accounted for by variation of their tertiary and quaternary structure, with for instance, equilibrium taking place between inactive monomers and active oligomers [3]. The scope of this chapter is to show how GPCRs may follow such a functional architecture and are submitted to regulation of their functionality not only by agonists and antagonists, but also by allosteric compounds.

Current knowledge about G protein-coupled receptors (GPCRs), allows description of this family of regulatory proteins within a multiple conformation allosteric scheme. On one hand, GPCRs undergo a spontaneous equilibration between inactive and active conformations that is regulated by agonists, antagonists and inverse agonists [4, 5]. On the other hand, it is evident from recent work that GPCRs couple to multiple intracellular pathways that are differentially regulated by distinct agonist molecules [6–10]. Finally, as for other allosteric proteins, co-activators or modulators of GPCR signaling have been identified [11, 12]. For some of them, characterization of their effects could be extended up to the animal level [13] or even to human testing [14].

G protein-coupled receptors comprise more than 800 members of which 375 proteins make up the group of non-olfactory receptors with more than 120 still being orphan proteins. All G protein-coupled receptors are regulatory proteins, i. e. they exist in different functional states (resting, active, desensitized etc.) which are regulated by pharmacological agents comprising endogenous ligands as well

as synthetic compounds [15] (see also Chapter 2). Given that they are located at the surface of cells, and that they regulate almost all physiological functions in higher organisms, G protein-coupled receptors have become the largest family of drug targets [12, 16, 17]. Until recently, useful GPCR-targeting pharmacological agents either used to decipher signaling pathways or to treat patients, were almost exclusively agonists or antagonists of the receptor protein. Detailed characterization of ligand binding to G protein-coupled receptors has led to descriptions of unconventional modes of action of agonists [18, 19] as well as antagonists [20], and to the discovery of so-called modulators which do not act by themselves and do not prevent agonist binding, but alter the response to it [12]. Such compounds that bind to sites that are topographically distinct from the agonist binding site and alter the conformation of the protein in a reversible manner, have been termed allosteric effectors [21]. Today, the word allosteric is most commonly used to describe the indirect action of molecules binding to a regulatory site that is distinct from the agonist binding site [22].

G protein-coupled receptors share structural and functional properties with, for instance, a common transmembrane organization with seven-helical domains and intracellular loops involved in the coupling of the receptor to signaling effectors. Due to amino acid substitutions among individual members of the family, G protein-coupled receptors exhibit a multiplicity of phenotypes manifested, among others, by a wide diversity of binding, signal transduction and possibly folding phenotypes. As a result of these variations, pharmacological specificities, kinetics of activation and desensitization, selectivity for intracellular effectors of the response, efficacy of coupling to signaling pathways, and possibly the number of active or inactive conformational states may vary significantly.

The most general description of the functional architecture of regulatory proteins, and of G protein-coupled receptors in particular, involves three major rules:

- Receptor molecules exist in several conformations that correspond to thermodynamically stable states with defined tertiary, and possibly quaternary, structure. Each conformation can be qualitatively described in structural terms (even if experimental evidence is not yet available) and functionally described as inactive, active or desensitized. In addition each state is characterized by its intrinsic affinity for the natural agonist or other ligands.

- The interconversion between any two conformational states occurs freely with an allosteric equilibrium constant. Ligands stabilize the conformations to which they bind with higher affinity.

- One given receptor, with defined primary, secondary and tertiary structure, has access to a unique set of conformational states, presumably further constrained by the membrane environment, coupling to effectors, and possibly by the fact that there may be more than one active or desensitized state.

Gene duplication, single nucleotide polymorphism, alternative splicing or species-specific mutations may alter any of the aforementioned properties of the regulatory protein. For instance, changing intrinsic binding properties may result in

an alteration of the pattern of conformations, of evoked or spontaneous biological activity, or of protein export, stability and trafficking.

6.2
Multiple Conformations and Signaling Pathways of G Protein-coupled Receptors

The most widely accepted model for GPCR activation is the extended ternary complex model (often referred to as the two-state model; references listed in [4, 5, 23]) according to which (1) receptor molecules interconvert between two conformations endowed with distinct structural, pharmacological and functional properties, and (2) receptors can spontaneously adopt an active conformation and couple to G proteins (reference listed in [23]), thereby accounting for basal response levels. Such spontaneous equilibrium between resting and active conformations has been successfully demonstrated for receptors that couple to the regulation of intracellular cAMP levels. The inhibition of basal cAMP production levels is brought about by a category of pharmacological agents, now referred to as inverse agonists [5, 24]. In the case of the 5HT4 receptor, the equilibrium constant between resting and active states was estimated to be around 0.1 to 0.2, i.e. about 15% of the receptor molecules spontaneously interconvert to the active state.

Although the two-state model aims to describe receptor properties within a minimal number of conformations, it is becoming increasingly clear that it no longer adequately mirrors the complexity of GPCR signaling.

A growing body of experimental evidence supports the idea that GPCRs exist in several conformations, among which multiple active states would mediate distinct signaling pathways through activation of, for instance, different hetero-trimeric G proteins. Examples of such dual coupling of receptors to signaling pathways include the dopamine receptors D1 (Gs and Go [25, 26]) and D5 (Gs and Gz [27]), the receptors for the parathyroid hormone (Gs, Gq and Gi [28]), for corticotropin-releasing hormone (Gs, Gi, Gq, Go and Gz [29]), for melanin-concentrating hormone (Gi and Go maybe Gq [30]), for vasoactive intestinal peptide (Gs and Gi [31]), for prostacyclin (Gs, Gi and Gq [32]), the adenosine A1 receptor (Go and Gi [33]), the beta2-adrenergic receptor (Gs and Gi [6]), the muscarinic m3 receptor (Gq and G12 [34]), the 5-HT4 receptor (Gi, Go and Gs [35]), and the endothelin subtype B receptor (Gi, Go and Gq [17, 36]). This does not only result from overexpression of receptors in heterologous expression systems, but also occurs under more physiological conditions where receptors are endogenously expressed [25, 31, 37].

Additional reported evidence in favor of multiple conformations of GPCRs include, among others, the fact that distinct agonists of opioid receptors differentially stimulate receptor phosphorylation and endocytosis [7], that the cholecystokinin receptor antagonist D-Trp-OPE leads to receptor internalization without promoting its phosphorylation [38], phosphorylation of the angiotensin receptor occurs in a conformation that differs from the active state [39], and finally, depending on the agonist used, NK2 tachykinin receptors can be repeatedly activated or not [40].

Theoretical descriptions of G protein-coupled receptors within a framework of multiple conformational states now include such diversity [10, 41–45].

6.2.1
Biophysical Approaches to Monitoring Conformational Changes of G Protein-coupled Receptors

Several approaches have been used to show that G protein-coupled receptors undergo conformational changes in response to agonist, as expected for allosteric proteins. In the examples reported below, conformational transitions were detected using fluorescence techniques, in particular by taking advantage of the physical process of fluorescence resonance energy transfer.

The first example reports real-time recording of ligand binding and parallel monitoring of intracellular responses in order to temporally correlate binding events with metabolic changes in the cell [9, 10, 42]. This approach, carried out on NK2 tachykinin receptors, makes use of a fluorescently labeled NK2 receptor (tagged with EGFP at its N-terminus) and of fluorescently labeled neurokinin A (NKA) derivatives chemically labeled with energy transfers acceptors of excited EGFP (bodipy or Texas red groups). Two agonists were used: one is full length neurokinin A that comprises 10 amino acids and is labeled at its N-terminus with Texas Red and the second is truncated NKA, NKA4−10, a naturally occurring agonist of NK2 receptors found in fluids from mid-gut cancers [46], that differs from NKA in its capacity to promote repeatable intracellular calcium responses upon successive receptor stimulations [9, 40], while full length NKA does not. Binding of the two fluorescent agonists is followed in a real-time manner as an energy transfer signal that results from close proximity of the donor and acceptor fluorophores when the receptor–ligand complex is formed (Fig. 6.1). The time course of fluorescent NKA4−10 binding to the NK2 receptor is mono-exponential while that of fluorescent NKA exhibits a rapid binding phase followed by a 10-times slower association process [10]. Detailed analysis of the binding kinetics shows that both agonists initially associate to the receptor in a diffusion-limited manner that tem-

Fig. 6.1 Dynamics of agonist binding to the NK2 tachykinin receptor. The cDNA encoding rat NK2 tachykinin receptor is fused with the EGFP encoding cDNA in order to lead to the expression of a fluorescent chimeric receptor upon transfection in HEK 293 cells. Neurokinin A is labeled at its amino terminus with Texas Red to yield the fluorescent compound referred to as TR1-NKA. Truncated NKA4−10 is synthesized with Val4 to Cys substitution in order to introduce the Texas Red fluorophore at a position equivalent to residue 7 from full length NKA. The resulting agonist is known as TR7-NKA4−10. Association of TR1-NKA and TR7-NKA4−10 is followed in real time as a decrease of EGFP fluorescence emission at 510 nm when the cell suspension is illuminated at 470 nm (right panels). Rapid binding of TR7-NKA4−10 and rapid and slow binding of TR1-NKA are simulated with a three-state model of the NK2 receptor in which the receptor exists in a resting (R0) conformation in equilibrium with two active states R1, coupled to calcium release, and R2 coupled to cAMP production [10, 42]. Isomerization constants between conformations are given beneath vertical arrows; intrinsic binding affinities for each conformation are given above horizontal arrows.
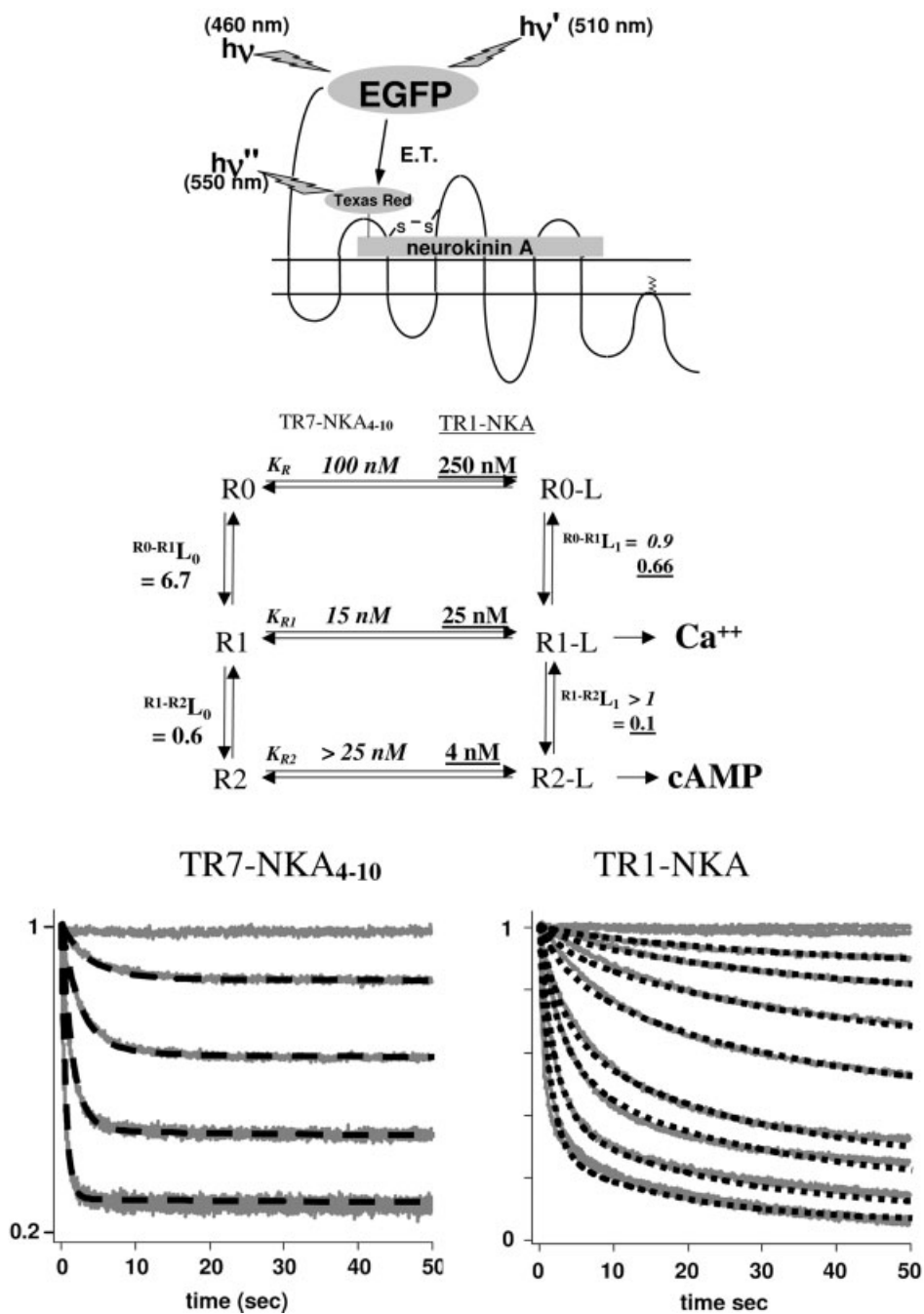
Fig. 6.1 (legend see page 102)

porally correlates with the onset of intracellular calcium elevation. The slow binding process that is observed only for full length NKA does not reflect a simple bimolecular association reaction but rather represent binding to a slowly appearing conformation that is not accessible to truncated NKA. The correlation of this slow binding with a metabolic change in the cell reveals that it matches the time course of cAMP formation in the cell, a signal that is detected in response to full length TR1-NKA but not to truncated TR7-NKA4–10 [10, 42]. The notion that the slowly appearing conformation that binds full length NKA which results from the interconversion of rapidly occupied receptor states, comes from the analysis of the dissociation of full length NKA from the receptor. Indeed, while NKA dissociates with a single exponential time course, when bound at equilibrium, the early initiation of dissociation following association (before equilibrium is reached) allows the detection of a rapid dissociation process, the amplitude of which decreases in favor of the slow dissociation process when the duration of association increases [10]. In other words, there is an intermediate state populated by full length TR1-NKA that progressively disappears as a result of an interconversion towards a stable state from which the agonist dissociates slowly.

Altogether, the binding and response data are interpreted within the framework of a three-state model (Fig. 6.1) in which the receptor exists in the following forms:

- a resting state R0 that is not coupled to any known intracellular response and binds agonists with low affinity (in the 200-nMnMOK? range),
- a rapidly stabilized active state R1 that couples to intracellular calcium signaling and binds both agonists with about 10-fold higher affinity (in the 20-nM range), and
- a slowly stabilized active state R2 that couples to cAMP formation, presumably through activation of Gs, and binds full length NKA with high affinity (in the 2-nM range) and truncated NKA with low affinity (KD > 20 nM).

Kinetic simulation [47] of the three-state model supports this interpretation and estimates relative abundances of each state, prior to agonist addition, that correspond to about 80% for the R0 conformation, 5–8% for the R1 state and 10–15% for the R2 state. Such values are in good agreement with experimentally determined values of spontaneous activity of the 5HT4 receptor [5] or H3 receptors [24] which exhibit about 10% of Gs coupled state in the absence of agonist. The rapid transition from the R0 to the R1 state was not detected in the study and the kinetic simulation of the binding traces accounts for this. Indeed, rapid binding to the active state R1 which triggers intracellular calcium elevation can be correctly simulated only because the rate of interconversion is faster than the rate of binding as has been detected in the concentration range explored.

A second energy transfer study [48] confirms this view. It reports the introduction of two fluorophores (CFP and YFP) within the same PTH or alpha2-adrenergic receptor molecule showing that activation of the receptor by PTH (PTH receptor) or clonidine (alpha2-adrenergic receptor) leads to increased CFP fluorescence and a concomitant decrease in YFP fluorescence intensities. The data thus show

that intracellular loop 3 that carries CFP and the C-terminal end of the receptor that carries YFP, become more distant from one another upon receptor activation by an agonist. The doubly-labeled receptors are biologically active, and their pharmacological profile follows that of unmodified receptors with agonists promoting variation and competitive antagonists inhibiting it. Although the authors analyze only rapid relaxation of the fluorescence signal as a single exponential relaxation, it is however evident from the reported data that the signal follows at least a bi-exponential time course reflecting, in the case of rapid relaxation, activation of the intracellular calcium response (adrenergic receptor) followed by an undetermined slower relaxation, the amplitude of which may represent up to 50% of the total fluorescence signal amplitude. This second relaxation likely reflects further stabilization of the PTH and adrenergic receptors into additional conformations of the receptor that might couple the receptor to other signaling pathways as previously proposed [6, 10, 49].

6.3
Allosteric Modulators of G Protein-coupled Receptors

A growing number of molecules interacting with GPCRs do not interact at the natural ligand (orthosteric) binding site. Rather, such compounds were discovered as molecules that modulate either positively or negatively the activity of a given GPCR or its interaction with conventional ligands such as agonists or antagonists [12, 50].

The most common way to describe allosteric ligands effects is to show that they modulate agonist or antagonist binding. Initial work in the field of GPCRs was mainly carried out with muscarinic and serotonin receptors [51–53]. Allosteric effectors of muscarinic receptors, mainly brucine, strychinine and gallamine inhibit dissociation of tritiated N-methyl scopolamine supporting the idea that they stabilize the receptor in a conformation of high affinity for antagonists [54]. However, the same modulators also potentiate binding of agonists and accordingly potentiate cellular responses to acetylcholine and carbachol [55, 56] supporting an unconventional mode of modulation.

Work done on other GPCRs has shown that allosteric effectors can be identified for almost all GPCRs, as has been illustrated for receptors of adrenaline, adenosine, chemokines, dopamine, GABA, glutamate, neurokinins, purines, and serotonin (e.g. [57–59]; reviewed in [12]).

In the case of the NK2 tachykinin receptor, a new allosteric modulator, LPI 805, was discovered as a dual modulator of NK2 receptor responses [60]. This compound speeds up dissociation of the natural agonist neurokinin A in a manner that is consistent with stabilization of a low affinity conformation which appears at the expense of a conformation exhibiting high affinity for neurokinin A (Fig. 6.2). Neurokinin A has been formerly shown to stabilize two active conformations of the NK2 tachykinin receptor [10, 42], linked to intracellular calcium release (low affinity active state) and cAMP production (active state exhibiting high
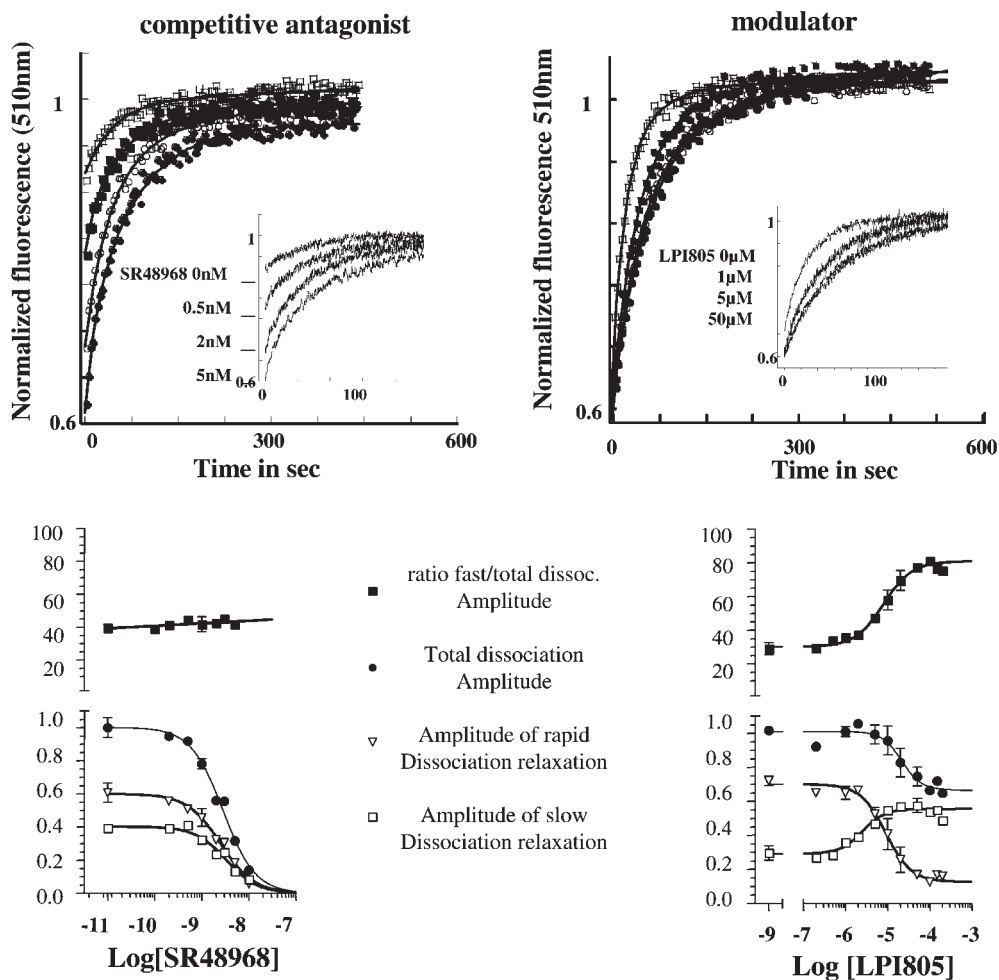
Fig. 6.2 Differential effects of the competitive antagonist SR 48968 and of the allosteric modulator LPI 805 on fluorescent neurokinin A dissociation. Left: dissociation of fluorescent neurokinin A, initiated by addition of a large excess of unlabeled neurokinin A, in the presence of varying concentrations of SR 48968 exhibits the same biphasic time course. Lower left graph shows that rapid and slow dissociation processes take place with identical relative amplitudes at all SR 48968 concentrations. Right: dissociation of fluorescent neurokinin A, initiated by addition of a large excess of unlabeled neurokinin A, in the presence of varying concentrations of LPI 805 exhibits the faster biphasic time course. Lower right panel shows that the amplitude of rapid dissociation increases at the expense of that of slow dissociation.
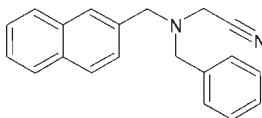
affinity for neurokinin A). Consistent with the effects of LPI805 on neurokinin A binding, calcium responses are potentiated by the allosteric effector while cAMP production is reduced or even abolished at high effector concentrations (Fig. 6.3). Modulation of GPCR function by allosteric modulators thus not only affects quantitative aspects of signaling, but also qualitatively affects GPCR function as they show a capacity for modifying the repertoire of cellular responses linked to GPCRs.

6.4
Where Do Allosteric Modulators Bind on GPCRs?

The most widely accepted model for modulation by allosteric effectors assumes that they bind on the receptor to a site that is completely distinct from the orthosteric site. This has been documented for nicotinic acetylcholine [61–63] and GABA A [64] receptors in particular. In the case of the G protein-coupled receptors, and depending on receptor subclasses, different modes of interactions may be used by the effector.

In the case of the NK2 receptor, the modulator LPI805 might bind to a site following the standard description for allosteric sites, i.e. occupancy of the allosteric site alters the equilibria between conformations and modulates agonist binding according to the intrinsic affinity of each conformation for the ligand. Neurokinin A indeed binds to the resting conformation of the NK2 receptor with a dissociation constant equal to 200–400 nM, to the active conformation 1 (linked to calcium release) with a KD equal to 20–40 nM and to the active conformation 2 (linked to cAMP production) with a KD equal to 2–5 nM [10]. By increasing the proportion of conformation active 1 versus active 2, the binding of LPI 805 leads to the apparent reduction of neurokinin A binding without otherwise affecting the rate of access of neurokinin to its site, nor changing the rates of dissociation from active 1 or active 2 conformations [60]. The site of LPI 805 thus likely lies on the receptor molecule at a topographical location distinct from that of the agonist (Fig. 6.4).

In the case of muscarinic receptors, allosteric modulation is commonly analyzed in terms of binding cooperativity according to the ternary complex model [52, 65, 66]. Depending on the receptor subtype and on the orthosteric ligand structure, the binding of a given allosteric molecule leads to increased, unchanged or decreased apparent affinity for the orthosteric ligand [67, 68]. In most cases, positive, neutral and negative cooperativity is associated with slowing down effects on orthosteric ligand dissociation which proceeds according to a mono-exponential decay. Magnitude and direction of the modulation of the affinity of a given receptor subtype is independent of the pharmacological nature (agonist or antagonist) of the primary ligand, an observation that is not in accord with an allosteric transition model between pre-existing conformational states [1, 62, 69].

It has also been shown that classical allosteric muscarinic modulators such as alcuronium and strychnine may exert intrinsic agonist-like effects (through M1, M2, M3 and M4 subtypes) that are not prevented by an antagonist [56]. They are
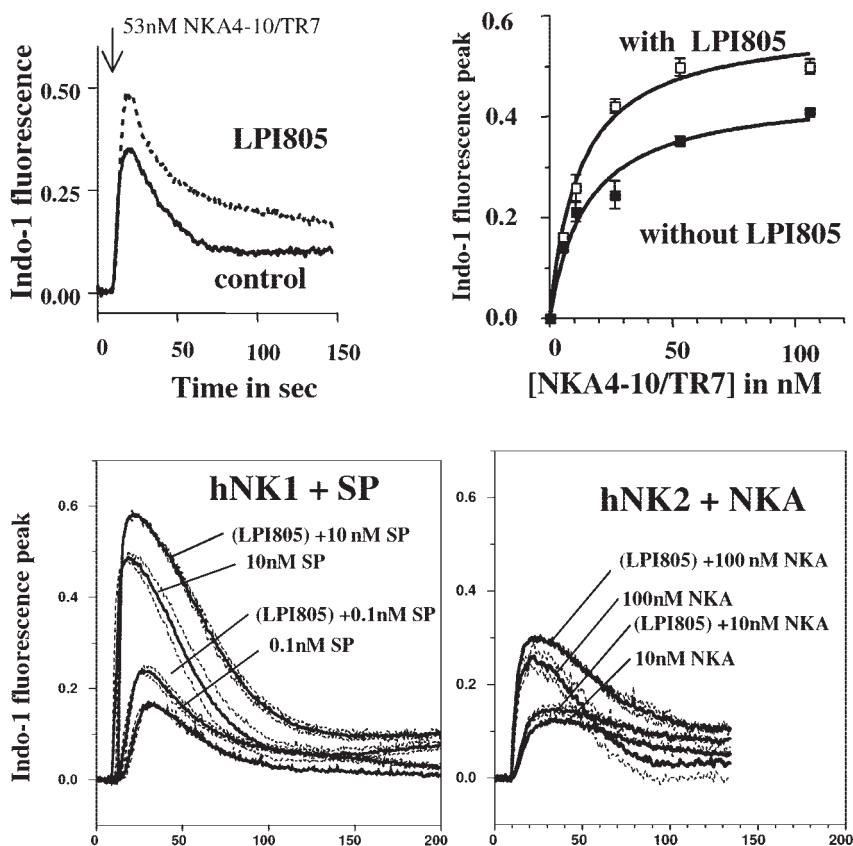
Calcium responses

EGFP-rNK2

Fig. 6.3 Actions of the allosteric effector LPI 805 on the NK2 receptor responses. The allosteric effector LPI 805 potentiates intracellular calcium responses of rat NK2 receptor, triggered by the partial agonist TR7-NKA4–10. Peak calcium responses are shown as a function of agonist concentration. LPI 805 also potentiates human NK1 receptor (activated by Substance P) and human NK2 receptor (activated by NKA) at different agonist concentrations. The allosteric effector LPI 805 non-competitively inhibits cAMP production. With regard to cAMP production dose–response relationships, LPI 805 leads to dose-dependent reduction of maximal response amplitude without affecting apparent response affinity. Maximal inhibition of cAMP response reaches 80 % at saturating concentrations and occurs with an apparent affinity (8 μM) identical to that of altered NKA dissociation rates. The effect of LPI 805 is distinct from the effect of a known competitive antagonist, H8565. H8565 does not modify the maximal cAMP response amplitude but alters the apparent affinity of the response in a dose-dependent manner.

cAMP responses
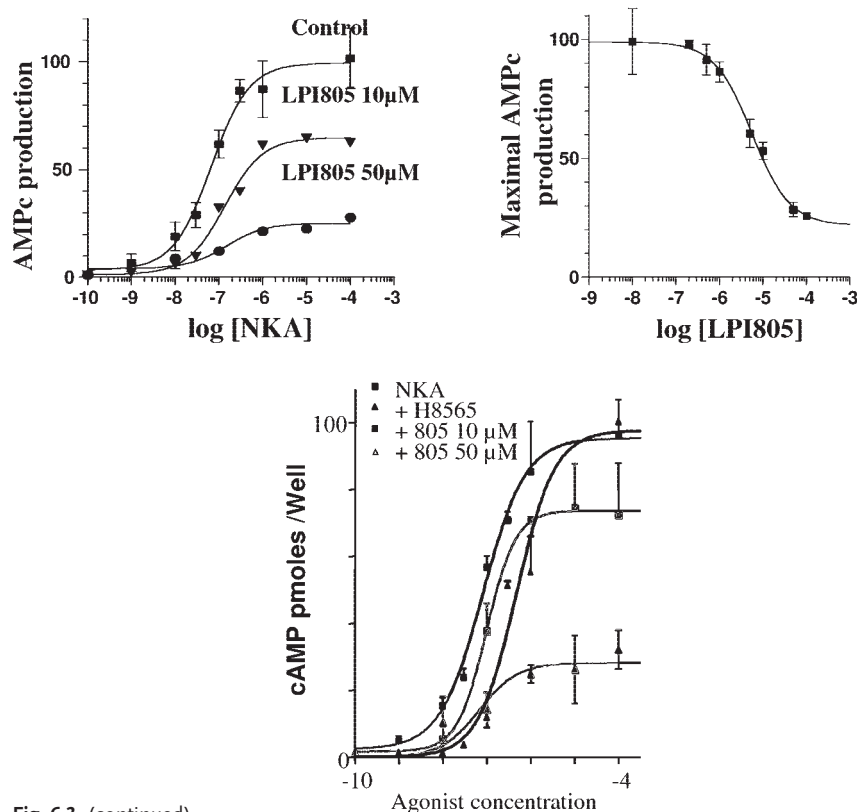
Fig. 6.3 (continued)

thus able, like allosteric agonists [69], to shift receptor equilibrium towards active conformations. This is at variance with a number of studies reporting, for the same modulators, increases in ^{3}H-antagonist (NMS) binding to M2 and M4 sub-types together with negative allosteric effects on acetylcholine binding [55, 56]. Most interestingly, alcuronium has recently been shown to display intrinsic in-verse agonist activity at M2 receptors and to suppress pilocarpine intrinsic efficacy to induce G protein activation [70]. This is a first example of a prototypical allos-teric modulator that appears to preferentially stabilize inactive conformations of the muscarinic M2 receptor.

There are several difficulties in defining a non-ambiguous allosteric modulation of muscarinic receptors and in evaluating its impact on equilibria between confor-mational and functional states of the receptors: (a) the allosteric site(s) is thought to be positioned above the orthosteric site in the extracellular regions of the mus-carinic receptor [71–74]. Thus, it may be considered, taking the example of alcuro-nium [75], that some modulators may bind in close proximity to the orthosteric site and sterically block access to (and/or departure from) this site (Fig. 6.4). As a
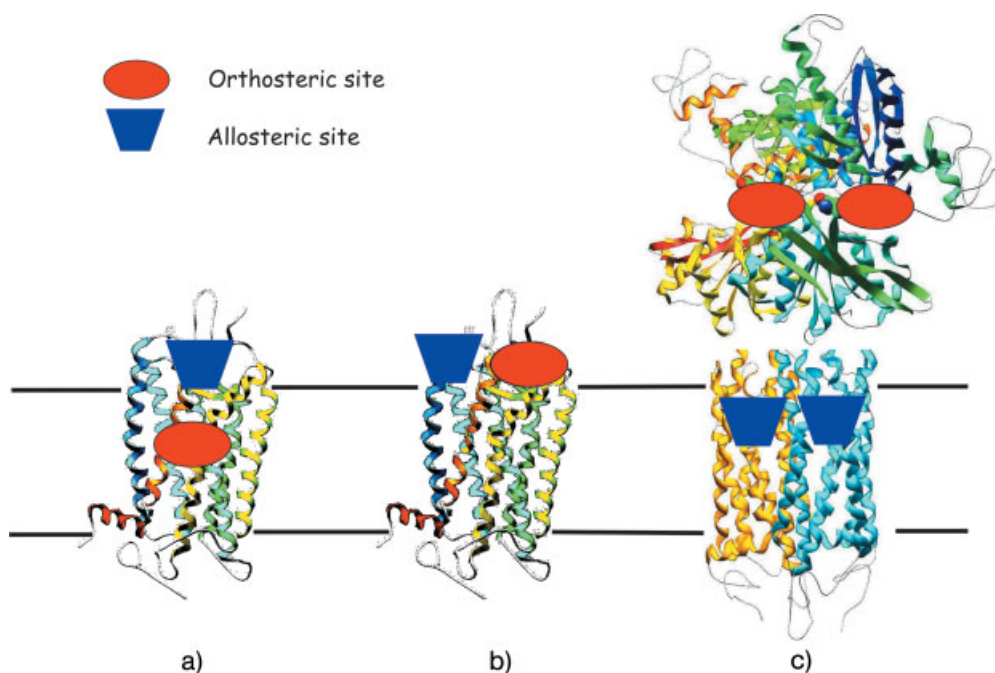
Fig. 6.4 Models of possible allosteric binding site locations on GPCRs. (a) The allosteric effector (blue lozenge) sterically hinders access and dissociation from the orthosteric site. (b) The allosteric and orthosteric sites are close to each other and ligands of one category of site bind without sterically affecting the binding of ligands to the other site. (c) Allosteric effectors of glutamate and related G protein-coupled receptors bind to the transmembrane domain of the protein, while orthosteric ligands bind to the large extracellular domain.

consequence, their macroscopic effect on receptor binding and function may reflect the contribution of steric hindrance and of true conformational changes of the orthosteric binding site; (b) experiments are usually performed on membrane preparations, under conditions which avoid the confounding effects related to G protein coupling [76] and which fit the ternary complex model; (c) most studies analyze variations in [^{3}H]-NMS dissociation kinetics, i.e. variations of an antagonist that may not discriminate between receptor affinity states; (d) direct evaluation of the effects of allosteric modulator candidates on ^{3}H-agonist (acetylcholine) binding properties is limited to the exploration of high affinity state receptor subpopulations. Indeed, high levels in non-specific binding preclude studies covering nanomolar to micromolar agonist concentrations; (e) fluorescent agonist probes are not available, rendering real-time monitoring of binding and functional events impossible.

Such profound discrepancies between binding and functional data strengthen the importance of the choice of a method to screen for new allosteric molecules and of the model to evaluate their mechanism of action.

Finally, an additional mode of interaction of allosteric effectors has been described for class III GPCRs such as glutamate, calcium or GABA B receptors, which possess a large extracellular domain to which the agonist and competitive antagonists bind. In these receptors, there is no evidence that the binding pocket which exists within the transmembrane portion of the receptor molecule is used under normal physiological conditions. This pocket can be occupied by small organic molecules that thus behave as allosteric modulators of this class of receptors (Fig. 6.4) [13, 58, 77]. The demonstration of such a mode of interaction has been confirmed experimentally using mutant glutamate receptors lacking the extracellular domain [58] and the calcium receptor [78, 79]. On truncated glutamate receptors, positive allosteric modulators of glutamate receptors behave as agonists.

6.5
Future Challenges for Allosteric Modulation of GPCRs

Allosteric modulators have been attracting increasing interest both from academic research and pharmaceutical companies. Since they bind to sites that are topographically distinct from the orthosteric site, the possibility of identifying compounds that are endowed with greater receptor subtype selectivity than are agonists, especially for receptor families that involve large numbers of subtypes, was being vigorously pursued. This issue had not been extensively studied however, and when carefully investigated, allosteric ligands were found to exhibit limited receptor specificity, as were other types of pharmacological agents [80].

Allosteric modulators are expected to show several advantages over classic drugs [12]. Indeed, as allosteric compounds do not exhibit significant activating/inhibiting activity in the absence of the endogenous ligand, their effect is mostly revealed under normal physiological conditions, i.e. when the agonist is released or secreted. They would thus not provoke sustained receptor activation, and therefore would not stimulate the desensitizing or downregulating processes that are traditionally triggered by agonists. It is thus expected that the risks associated with clinical overdose will be reduced and that allosteric effectors, by modulating the ratio of distinct active conformations may produce a large variety of combined qualitative and quantitative effects which would lead to an extensive array of original pharmacological properties (for a review see [69]). It is however clear that such an approach will only be pertinent if the defect is not a lack of transmitter production or release. Finally, even negative allosteric effectors may represent advantageous alternatives to traditional competitive antagonists in terms of therapeutic use as the fine-tuning of therapeutic dosages is often difficult to achieve because of inter-individual differences. By setting the maximal level of receptor activation, as a result of fixing the equilibrium between active and inactive states, negative allosteric modulators could thus also be prescribed in an attempt to reduce the risk of dosage side-effects.

Despite considerable research to show that natural and synthetic molecules can modulate the function of GPCRs, and thus behave as allosteric effectors, there is

still limited evidence to indicate that such molecules do indeed exert their effects through direct interaction with the receptor protein [68]. Most allosteric compounds do indeed exhibit moderate affinity for their target receptors which makes direct binding experiments difficult as a result of the elevated levels of non-specific binding. Although site-directed mutagenesis experiments [71, 73, 78, 79] convincingly show binding or functional alterations, they may also affect either intrinsic binding affinity directly or indirectly as a result of the modified equilibrium between conformations. They are thus difficult to interpret in structural terms and research efforts will therefore need to be directed towards binding experiments, affinity or photoaffinity labeling and other structural approaches.

References

1 MONOD, J., J. WYMAN, and J.P. CHANGEUX. *J Mol Biol* **1965**; *12*: 88–118.

2 PERUTZ, M.F. *Q Rev Biophys* **1989**; *22*: 139–237.

3 QIU, Y., et al. *J Biol Chem* **2000**; *275*: 13517–13528.

4 COTECCHIA, S., et al. *Proc Natl Acad Sci U S A* **1990**; *87*: 2896–2900.

5 CLAEYSEN, S., et al. *EMBO Rep* **2001**; *2*: 61–67.

6 DAAKA, Y., L.M. LUTTRELL, and R.J. LEFKOWITZ. *Nature* **1997**; *390*: 88–91.

7 WHISTLER, J.L., et al. *Neuron* **1999**; *23*: 737–746.

8 SAGAN, S., et al. *J Pharmacol Exp Ther* **1996**; *276*: 1039–1048.

9 VOLLMER, J.Y., et al. *J Biol Chem* **1999**; *274*: 37915–37922.

10 PALANCHE, T., et al. *J Biol Chem* **2001**; *276*: 34853–34861.

11 CHRISTOPOULOS, A. and T. KENAKIN. *Pharmacol Rev* **2002**; *54*: 323–374.

12 CHRISTOPOULOS, A. *Nat Rev Drug Discov* **2002**; *1*: 198–210.

13 URWYLER, S., et al. *Mol Pharmacol* **2001**; *60*: 963–971.

14 CONIGRAVE, A.D., S.J. QUINN, and E.M. BROWN. *Trends Pharmacol Sci* **2000**; *21*: 401–407.

15 PIERCE, K.L., R.T. PREMONT, and R.J. LEFKOWITZ. *Nat Rev Mol Cell Biol* **2002**; *3*: 639–650.

16 WILSON, S., et al. *Br J Pharmacol* **1998**; *125*: 1387–1392.

17 SHAABAN, S. and B. BENTON. *Curr Opin Drug Discov Devel* **2001**; *4*: 535–547.

18 SPALDING, T.A., et al. *Mol Pharmacol* **2002**; *61*: 1297–1302.

19 SUR, C., et al. *Proc Natl Acad Sci USA* **2003**; *100*: 13674–13679.

20 FILLION, G., et al. *Adv Exp Med Biol* **1991**; *287*: 165–176.

21 JACOB, F. and J. MONOD. *J Mol Biol* **1961**; *3*: 318–356.

22 CHANGEUX, J.P. and S.J. EDELSTEIN. *Neuron* **1998**; *21*: 959–980.

23 GETHER, U. *Endocr Rev* **2000**; *21*: 90–113.

24 MORISSET, S., et al. *Nature* **2000**; *408*: 860–864.

25 KIMURA, K., B.H. WHITE, and A. SIDHU. *J Biol Chem* **1995**; *270*: 14672–14678.

26 LEZCANO, N., et al. *Science* **2000**; *287*: 1660–1664.

27 SIDHU, A. *Mol Neurobiol* **1998**; *16*: 125–134.

28 SCHWINDINGER, W.F., et al. *Endocrine* **1998**; *8*: 201–209.

29 GRAMMATOPOULOS, D.K., et al. *J Neurochem* **2001**; *76*: 509–519.

30 HAWES, B.E., et al. *Endocrinology* **2000**; *141*: 4524–4532.

31 LUO, X., et al. *J Biol Chem* **1999**; *274*: 17684–17690.

32 LAWLER, O.A., S.M. MIGGIN, and B.T. KINSELLA. *J Biol Chem* **2001**; *276*: 33596–33607.

33 LORENZEN, A., H. LANG, and U. SCHWABE. *Biochem Pharmacol* **1998**; *56*: 1287–1293.

34 RUMENAPP, U., et al. *J Biol Chem* **2001**; *276*: 2474–2479.

35 Pindon, A., et al. *Mol Pharmacol* **2002**; *61*: 85–96.

36 Fuchs, S., et al. *Mol Med* **2001**; *7*: 115–124.

37 Andrade, R. *Neuron* **1993**; *10*: 83–88.

38 Roettger, B.F., et al. *Mol Pharmacol* **1997**; *51*: 357–362.

39 Thomas, W.G., et al. *J Biol Chem* **2000**; *275*: 2893–2900.

40 Nemeth, K. and A. Chollet. *J Biol Chem* **1995**; *270*: 27601–27605.

41 Leff, P., et al. *Trends Pharmacol Sci* **1997**; *18*: 355–362.

42 Lecat, S., et al. *J Biol Chem* **2002**; *277*: 42034–42048.

43 Kobilka, B.K. and U. Gether. *Methods Enzymol* **2002**; *343*: 170–182.

44 Kobilka, B.K. *J Pept Res* **2002**; *60*: 317–321.

45 Kenakin, T. *Receptors Channels* **2004**; *10*: 51–60.

46 Theodorsson-Norheim, E., et al. *Eur J Biochem* **1987**; *166*: 693–697.

47 Edelstein, S.J., Schaad, O., Henry, E., Bertrand, D., Changeux, J.P., *Biol Cyber* **1996**; *75*: 361–379.

48 Vilardaga, J.P., et al. *Nat Biotechnol* **2003**; *21*: 807–812.

49 DeFea, K.A., et al. *Proc Natl Acad Sci USA* **2000**; *97*: 11086–11091.

50 Rees, S., D. Morrow, and T. Kenakin. *Receptors Channels* **2002**; *8*: 261–268.

51 Massot, O., et al. *Mol Pharmacol* **1996**; *50*: 752–762.

52 Tucek, S. and J. Proska. *Trends Pharmacol Sci* **1995**; *16*: 205–212.

53 Birdsall, N.J., et al. *Mol Pharmacol* **1999**; *55*: 778–786.

54 Lazareno, S., et al. *Life Sci* **1999**; *64*: 519–526.

55 Lazareno, S., et al. *Mol Pharmacol* **1998**; *53*: 573–589.

56 Jakubik, J., et al. *Mol Pharmacol* **1997**; *52*: 172–179.

57 Gao, Z.G., et al. *Biochem Pharmacol* **2003**; *65*: 525–534.

58 Goudet, C., et al. *Proc Natl Acad Sci USA* **2004**; *101*: 378–383.

59 Thomas, E.A., et al. *Proc Natl Acad Sci USA* **1997**; *94*: 14115–14119.

60 Galzi, J.L., Hibert, M., Bourguignon, J.J., Maillet, E., Publication no. 2 840 993, Bulletin Officiel de la Propriété Industrielle no. 05/14 du 08/04/05, 17 june 2002, france.

61 Heidmann, T., R.E. Oswald, and J.P. Changeux. *Biochemistry* **1983**; *22*: 3112–3127.

62 Galzi, J.L., S.J. Edelstein, and J. Changeux. *Proc Natl Acad Sci USA* **1996**; *93*: 1853–1858.

63 Krause, R.M., et al. *Mol Pharmacol* **1998**; *53*: 283–294.

64 Galzi, J. and J. Changeux. *Currt Opin Struct Biol* **1994**; *4*: 554–565.

65 Stockton, J.M., et al. *Mol Pharmacol* **1983**; *23*: 551–557.

66 Ehlert, F.J. *Mol Pharmacol* **1988**; *33*: 187–194.

67 Christopoulos, A., et al. *Mol Pharmacol* **1998**; *53*: 1120–1130.

68 Trankle, C., et al. *Mol Pharmacol* **1998**; *54*: 139–145.

69 Hall, D.A. *Mol Pharmacol* **2000**; *58*: 1412–1423.

70 Zahn, K., et al. *J Pharmacol Exp Ther* **2002**; *301*: 720–728.

71 Matsui, H., S. Lazareno, and N.J. Birdsall. *Mol Pharmacol* **1995**; *47*: 88–98.

72 Gnagey, A.L., M. Seidenberg, and J. Ellis. *Mol Pharmacol* **1999**; *56*: 1245–1253.

73 Voigtlander, U., et al. *Mol Pharmacol* **2003**; *64*: 21–31.

74 Jakubik, J., A. Krejci, and V. Dolezal. *J Pharmacol Exp Ther* **2005**; *313*: 688–96.

75 Proska, J. and S. Tucek. *Mol Pharmacol* **1994**; *45*: 709–717.

76 Lazareno, S. *Methods Mol Biol* **2004**; *259*: 29–46.

77 Conigrave, A.D., S.J. Quinn, and E.M. Brown. *Proc Natl Acad Sci USA* **2000**; *97*: 4814–4819.

78 Petrel, C., et al. *J Biol Chem* **2004**; *279*: 18990–18997.

79 Petrel, C., et al. *J Biol Chem* **2003**; *278*: 49487–49494.

80 Bradley, K., E. Rowan, and A. Harvey. *Toxicon* **2001**; *39*: 581–587.

7
Chemogenomics Approaches to Ligand Design

Thomas Klabunde

7.1
Introduction to Chemogenomics: Similar Receptors Bind Similar Ligands

Among the 100 top-selling drugs 25 act on G protein-coupled receptors (GPCRs) [1]. The current estimate of the number of GPCRs in the human genome (excluding olfactory receptors) is about 400 [2, 3]. Currently, the drugs available on the market address only 30 GPCRs, which represent a small fraction of the GPCR target space. Thus there are approximately 370 further GPCRs which may be promising drug targets and may provide excellent opportunities for successful drug discovery programs. Therefore, most pharmaceutical companies invest heavily in this target family following different approaches to identify novel GPCR hit series. High-throughput screening (HTS) lead-finding approaches are often complemented with rational chemogenomics approaches to ligand design; these are highlighted within this chapter.

Several recent review articles have been published on "chemogenomics" with the aim of providing a clear definition of this novel approach to drug discovery [4–7]. A shift in pharmaceutical research from traditional target-specific case-by-case studies to a cross-target view might be considered as the common underlying theme. Following the "chemogenomics" concept, targets are no longer viewed as individual and single entities but grouped into sets of related proteins or target families (e.g. kinases, GPCRs) that are systematically explored. Today, many pharmaceutical companies have adapted their traditional organizational structure and are following a chemogenomics approach to increase the efficiency of modern drug discovery [5, 8].

Within the chemogenomics drug discovery scenario compounds are profiled against several targets and not only tested against single targets. Profiling of novel leads (or even of known drugs) [9] for a given target on related molecular targets and selective chemical optimization of "side affinities" can provide additional attractive lead series [10]. In addition, novel chemoinformatic and bioinformatics tools have been developed to systematically investigate drug–target interactions and to derive insights into privileged chemical structural motifs which are corre-

lated with a target family-wide commonality in ligand recognition [1, 11–13]. One main resulting value of the chemogenomics approach is the translation of these ligand recognition insights into the design and synthesis of chemical libraries to accelerate lead finding and target validation.

The paradigm which underlies these chemogenomics approaches to ligand design can be stated as "similar receptors bind similar ligands" [14]. This implies that for a novel GPCR the information obtained from known ligands for a related GPCR, can serve as a starting point for lead finding. Similarity, however, can be defined either by the scaffold or by the presence of identical fragments among the chemical structures being compared (Fig. 7.1). In addition, there are several other methods or descriptors of comparison and similarity metrics with which to compare chemical structures such as 2D fingerprints, 3D-pharmacophores, BCUT descriptors [15], "feature trees" [16], and CATs descriptors [17], which might all be relevant when defining similarity of biologically active molecules [18]. These descriptors can be used to support the selection of screening collections and the design of focused libraries directed against the GPCR target family (see Section 7.2).

Also the level of receptor similarity can be defined in different ways (Fig. 7.1). Using a coarse classification level proteins belonging to the same target family or
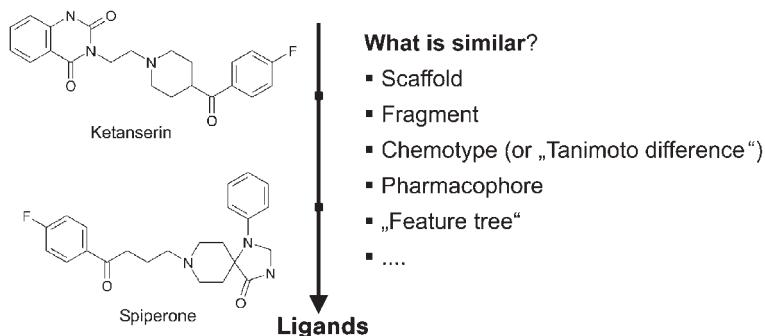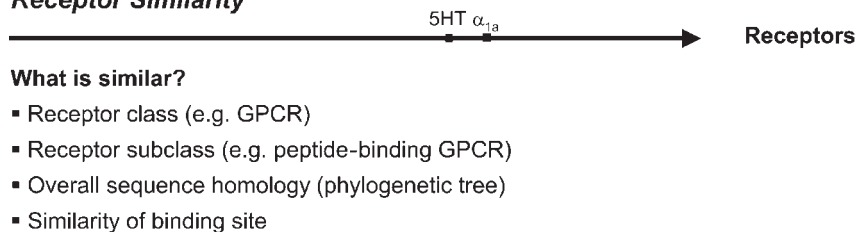
Fig. 7.1 Following the chemogenomics paradigm „Similar receptors bind similar ligands", the similarity of ligands and receptors can be defined using different methods and descriptors.

class, e.g. the family of G-protein coupled receptors (GPCRs) can be considered as similar. Although GPCRs all have seven-transmembrane helices and translate an extracellular signal into an intracellular response mediated by G-proteins (explaining the origin of their name), there is a great diversity of ligands for GPCRs. These can be small molecules like biogenic amines, amino acids, adenosine, lipids, and small peptides but proteins such as chemokines can also function as ligands. A better classification level at higher "resolution" (especially in the context of the design of libraries and screening sets for GPCRs) is thus defining two receptors as being similar if they bind the same class of ligands, e.g. peptides. This defines different GPCR subclasses such as chemokine receptors, peptide-binding GPCRs or purinergic GPCRs. Another similarity metric of receptors can be derived from the overall sequence similarity of the receptors. Finally, a further relevant viewpoint for a chemogenomics-driven classification approach would be the comparison of two receptors based on the similarity of their putative ligand binding sites [19], as this would be the best indication that two receptors bind similar ligands. However, there is insufficient reliable 3D structural information available regarding the GPCRs. Thus, insights into drug–receptor interactions have to be derived either from molecular recognition experiments (e.g. receptor labeling or site-directed mutagenesis) or from extensive profiling of compound series against sets of receptors and subsequent data analysis by computational tools (see Section 7.3).

7.2
Focused Libraries and Screening Collections Directed Against GPCRs

Due to the distinguished past of GPCR research there is a large number of currently marketed drugs acting at GPCRs [1]. In addition, for each of these marketed drugs there are hundreds of compounds that, whilst being potent ligands of GPCRs, have never been marketed. Following the chemogenomics paradigm this wealth of medicinal chemistry knowledge covering structural features of small-molecule GPCR ligands is an excellent source to enhance the discovery of new lead compounds against novel GPCR targets. Nowadays, databases covering ligand information and biological data are offered by several public and commercial providers [20]. For many pharmaceutical companies these databases provide the reference set which can be used to identify "GPCR-ligand-like" compounds within the companies' legacy compound collection by virtual screening and database searching. In addition, several molecular and pharmacophore descriptors have successfully been applied not only to define molecular similarity and to support the compilation of screening sets, but also to guide the design and synthesis of GPCR-directed libraries. It is beyond the scope (and not the purpose) of this chapter to provide an overview of all virtual screening, library design and synthesis studies targeting a single GPCR. The following sections seek to highlight those studies on libraries or screening sets designed to address the entire GPCR family or a specific GPCR subfamily.

7.2.1
Physicochemical Property-based Selection of GPCR Screening Sets

Researchers at Chemical Diversity Lab, Inc. have used a scoring scheme based on physicochemical properties for the classification of "GPCR-ligand-like" and "non-GPCR-ligand-like" compounds [21]. The methodology is a valuable tool to aid in the selection and prioritization of potential GPCR ligands from large collections and to guide the synthesis of novel GPCR-directed libraries. Using 5736 known GPCR ligands and 7506 non-GPCR ligands and eight physicochemical property descriptors (molecular weight, clogP at pH 7.4, number of H-bond donors, number of H-bond acceptors, number of rotatable bonds, solubility in water at pH 7.4 and fraction dose absorbed) a neural network model has been developed. Classification of randomly selected compound sets allowed the correct prediction of up to 92% of the GPCR ligands and 93% of the non-GPCR ligands within the training set. According to the authors the neural network model has also been applied to the generation of a GPCR-ligand-like focused library. By scoring of the companies CombiLab collection 30,000 compounds were selected and compiled into a GPCR-focused collection. No data concerning the biological testing of this collection has yet been disclosed. The authors also noted that the described procedure is amenable to virtual GPCR-focused combinatorial libraries and hence can guide the design and synthesis of GPCR-directed scaffold libraries.

7.2.2
Pharmacophore and Molecular Descriptors for GPCR Directed Libraries

In addition to the property-based classification and design methods presented in Section 7.2.1 several molecular and pharmacophore descriptors have been successfully applied to define molecular similarity and to support the compilation of screening sets for the design and synthesis of GPCR-directed libraries. Mason and co-workers have described a four-point pharmacophore method generating pharmacophore "keys" or "fingerprints" as a new measure of pharmacophore similarity and diversity: up to seven features and 15 distance ranges are considered, giving up to 350 million four-point 3D pharmacophores per molecule [22]. The authors were able to identify GPCR-specific pharmacophores and to apply the pharmacophore key to the design of GPCR-focused libraries.

A similar approach has recently been taken by Lamb and colleagues. In their article they describe a computational procedure to design a GPCR-directed screening library using pharmacophore descriptors of known GPCR ligands [23]. A reference set of 2785 known "drug-like" GPCR ligands was extracted from the MDL Drug Data Report database (MDDR). Three- and four-point pharmacophore descriptors were generated considering different chemical features within 25 distance bins. For each molecule of the training set a conformational model was generated using an in-house tool and the presence or absence of particular pharmacophores was recorded in a bit string („molecular signature"). Pharmacophores contained in more than 10 molecules were used to define the "GPCR drug space".

Subsequent iterative analysis allowed for the prioritization of scaffolds and sublibraries for inclusion into the final screening library, consisting of approximately 14,000 compounds. Screening of this library against the µ-opioid receptor provided 149 molecules around 10 different scaffold templates with an IC_{50} of at least 10 µM (see Fig. 7.2). The authors note that the gene family-target libraries which have been designed can provide an effective strategy for tackling multiple related targets simultaneously.
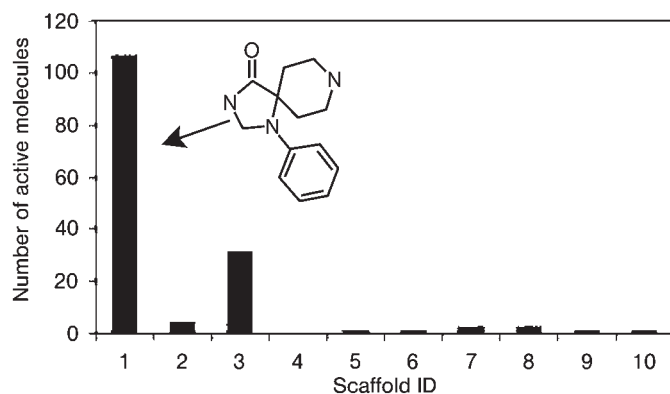
Fig. 7.2 Screening results of the GPCR-directed library tested against the µ-opioid receptor. The number of active molecules with an IC_{50} of less than 10 µM is plotted as a function of chemical scaffold assignment. (Reprinted from [23]; copyright 2004, with permission from Elsevier).

Pearlman and his colleagues have described the molecular BCUT descriptor, which combines physicochemical properties relevant to ligand-receptor binding with topological or distance information [15]. The BCUT descriptors have been utilized for a variety of purposes including reagent selection for chemically diverse libraries. The BCUT descriptors have been employed as input descriptors in a consensus neural network approach by Manallack et al. to distinguish compounds targeted towards specific gene families, including biogenic amine- and peptide-binding GPCRs [24]. The training sets (GPCR amine data, GPCR peptide data, random sets) were extracted from the MDDR database. The resulting neural network models correctly classified 81% of the ligands of biogenic amine-binding GPCRs and 87% of the ligands of peptide-binding GPCRs. The authors also mention that the consensus networks have been applied to the purchasing of compounds from third party suppliers for biological testing.

More recently researchers at Neurocrine have used the BCUT descriptors to define a region of chemical property space addressing a subset of GPCRs, which are activated by positively charged peptides (GPCR-PA$^+$) [25]. They defined the "drug space" by using 81,560 drugs and drug-like molecules for which BCUT metrics were calculated. The examination of the BCUT metrics supported the definition of

a five-dimensional drug space (H-bond donor and acceptor, two metrics of polarizability, charge). Each of the axes defining the "drug space" was divided into 10 bins resulting in the partitioning of the entire drug space into 100,000 cells. Compounds known to be active against the GPCR-PA$^+$ target subset (dataset of 630 molecules active against 40 GPCRs) were projected into the cell-divided drug space. The projection showed that the GPCR-PA$^+$ ligands including neighboring cells, occupy a subspace of only 6% of the entire drug space. Subsequently a virtual collection of around 19 scaffolds comprising 9,025,685 compounds was projected into the five-dimensional BCUT metrics space in order to evaluate their location relative to the defined GPCR-PA$^+$ space. The analysis supported the utilization of seven templates for library production and a test set of 2025 compounds was synthesized. The set was screened at a concentration of 15 μM using radioligand displacement assays against three members of the target GPCR subfamily: the melanocortin-4 receptor (MC4), the melanocyte concentrating hormone receptor (MCH) and the gonadrotropin-releasing hormone receptor (GnRH). In parallel a set of 2024 compounds randomly selected from Neurocrine's corporate screening collection was tested as a control library. Hit enrichment rates (a hit was defined as a compound displacing 50% or more of the specifically bound radioligand in duplicate tests) were found to range from 4.5-fold (GnRH-R) to 61-fold (MC4-R) when compared with the random set, revealing significantly higher numbers of hits identified in the designed set.

7.2.3
Privileged-fragment-based GPCR-directed Libraries

Small molecule ligands of diverse GPCRs often share common structural motifs, so-called "privileged structures". Originally Evans et al. introduced the term "privileged structure" for benzodiazepines, which are found in several types of CNS agents and ligands of ion channels and GPCRs [26]. Further examples of GPCR privileged substructures are biphenyl, 1,1-diphenylmethane, 4-arylpiperidines and piperazines, as well as spiro versions of the latter (see Fig. 7.3).

Several libraries based on privileged scaffolds or the incorporation of privileged fragments as building blocks have been synthesized as one strategy for the generation of focused libraries; examples for benzodiazepines, dihydropyridines, 4-phenylpiperidine and 1,1-diphenyl units have been described [27–29]. There are excellent reviews covering privileged structures [30, 31] and there is also a chapter in this book dedicated to the topic. The inherent promiscuity of privileged structures can also be a drawback, as selectivity might become an issue. However, ample evidence exists that small changes in and around the core can cause dramatic activity differences and result in useful selectivity [32, 33].
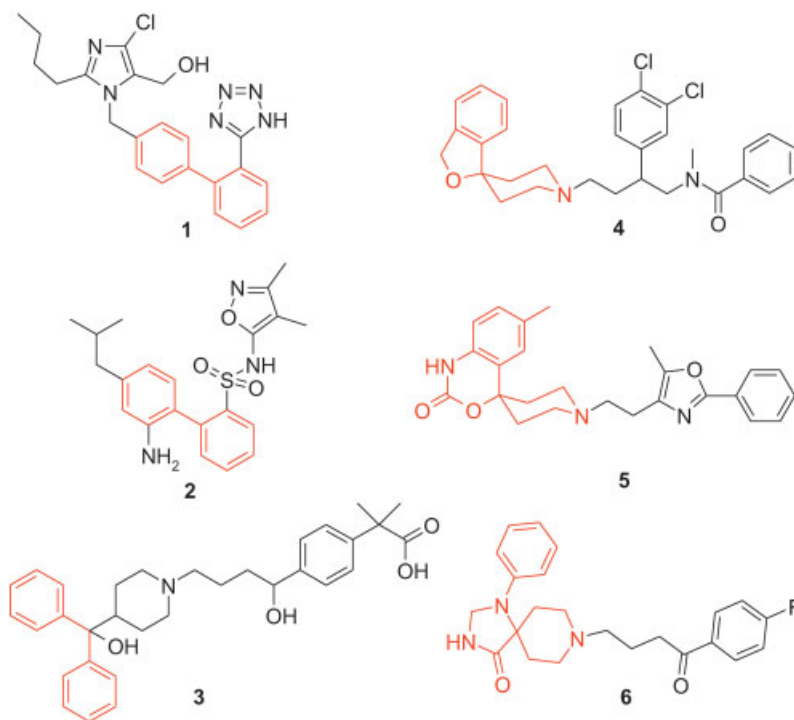
Fig. 7.3 GPCR-privileged substructures: examples of GPCR ligands sharing the biphenyl or diphenylmethyl (1–3) and spiropiperidine (4–6) moiety are shown [1]. **1** Angiotensin II type 1 antagonist (losartan); **2** endothelin-A antagonist; **3** histamine-H1 antagonist (fexofenadine); **4** neurokinin-NK$_2$ antagonist; **5** monocyte chemoattracant-1 (MCP-1)/CCR2 antagonist; **6** dopamine antagonist (spiperone).

7.2.4
GPCR Collection and Subfamily-directed Library Design

Within our company we are also following a rational chemogenomics approach for GPCR lead-finding which has already provided novel chemical series for several GPCR targets (e.g. P2Y$_{12}$ receptor, adenosine A1 receptor, chemokine CCR1 receptor). Our GPCR screening collection consists of a set of compounds with structural similarity to known GPCR ligands identified by virtual screening from various companies' compound collections (see Fig. 7.4) [34]. As a reference set of GPCR ligands defining the query compounds we used the MDDR (using compounds defined as GPCR ligands), the Aureus GPCR ligand database, and structural data from internal GPCR hits and leads. The MDDR covers patent literature, journals, meetings, and congresses and currently contains over 141,000 biologically relevant compounds and well-defined derivatives such as drugs launched or in the development phase [35]. The database compiled by Aureus Pharma focuses on GPCR ligands and covers all biological data including detailed information on
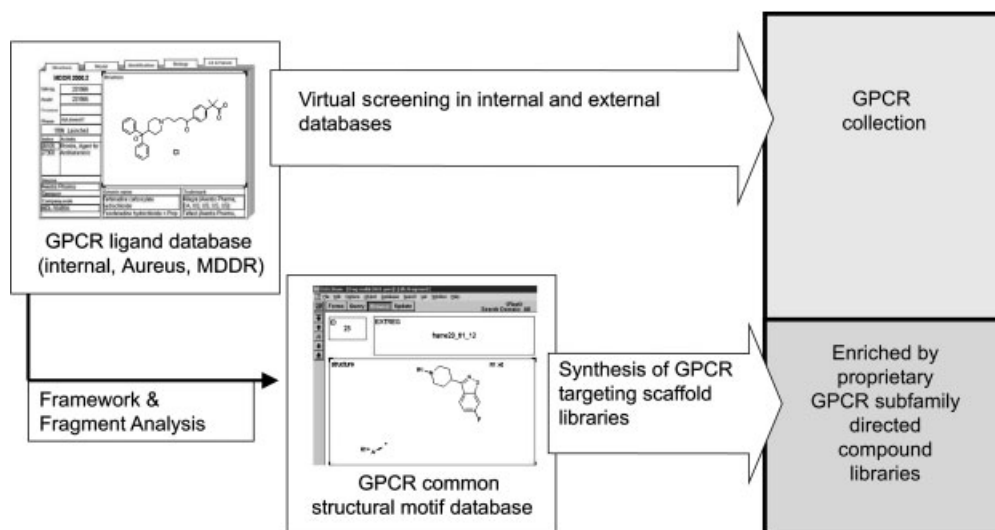
Fig. 7.4 Design process for a compound collection targeting GPCRs [34]. Details are given in the text.

experimental conditions [36]. Entries from this GPCR ligand database were taken as probes to perform similarity searches in internal and vendor compound collections using molecular 2D descriptors and feature tree descriptors as similarity indices. Compounds with structural similarity to known GPCR ligands were compiled as the basic set of the companies' GPCR compound collection. This basic collection targeting the complete GPCR family is supplemented by proprietary scaffold libraries (as shown in Fig. 7.4 lower part), which are designed and synthesized to address specific GPCR subfamilies (e.g. chemokine receptors, purinergic GPCRs). This target subfamily-directed library design and synthesis approach is illustrated below for the class of purinergic GPCRs.

The family of purinergic GPCRs represents a group of validated drug targets within a highly competitive field of GPCR research. It can be divided into P1 (binding nucleosides) and P2Y (binding nucleotides) receptors. Agonists and antagonists for these receptors have or are thought to have several therapeutic applications. Examples of drugs targeting this GPCR subfamily are doxofyline, an antagonist of the adenosine A1 receptor, used as bronchodilator, and clopidogrel, an antagonist of the $P2Y_{12}$ receptor, a platelet aggregation inhibitor prescribed for thrombosis and for the prevention of stroke. In order to identify novel and patentable leads for receptors of this group of GPCRs we have designed and synthesized proprietary scaffold libraries (Fig. 7.5 a) [34]. Using known ligands of the P1 and P2Y family pharmacophore models were generated by describing the key chemical elements required for binding to the P1 and P2Y family. Figure 7.5 a shows an example of the P2Y antagonist pharmacophore mapped onto AZD-6140, a known $P2Y_{12}$ antagonist. The 3D pharmacophore information was translated into 2D "de-
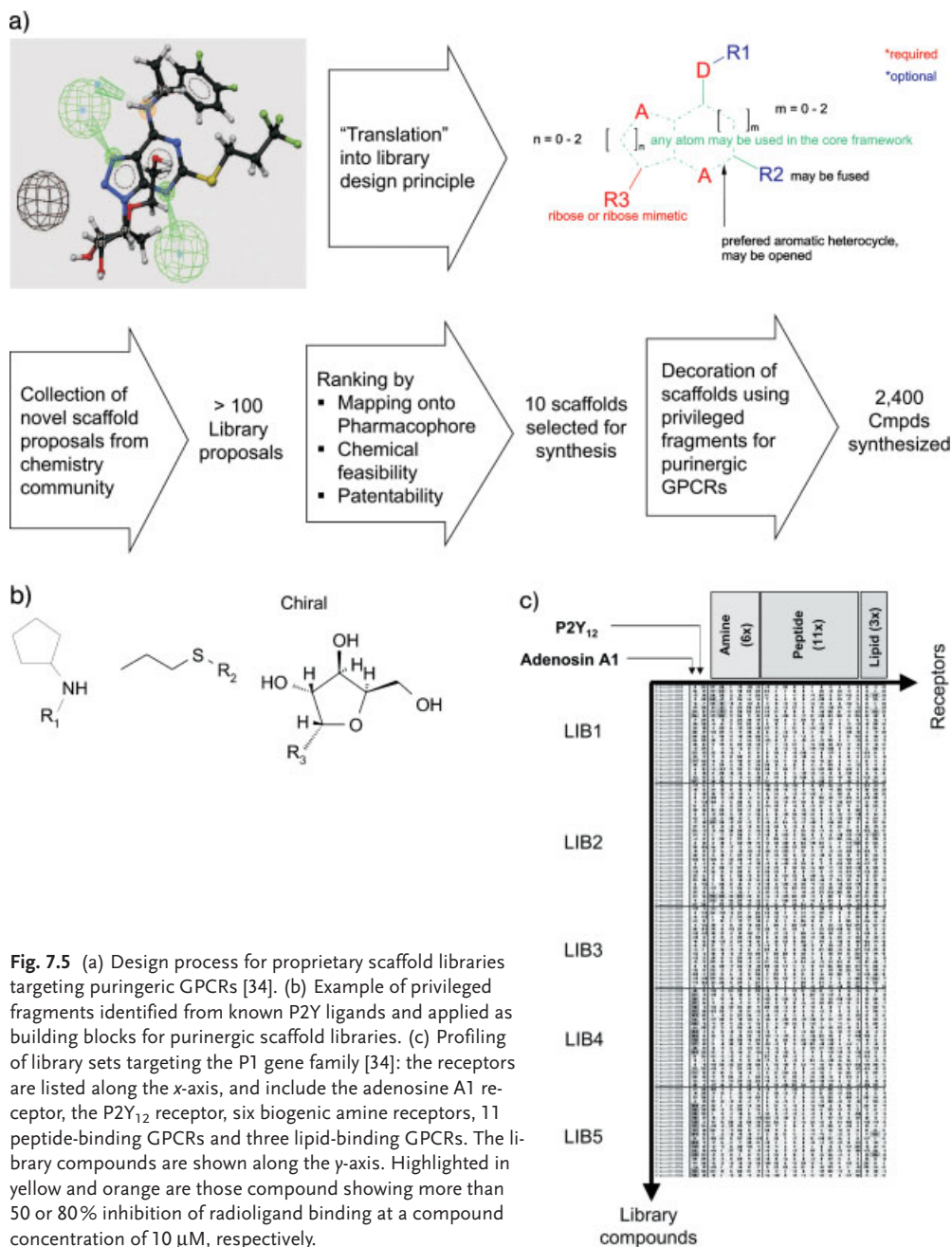
a)

"Translation" into library design principle

*required
*optional

R1
D
A
$n = 0 - 2$ [ ]$_n$ any atom may be used in the core framework
$m = 0 - 2$ [ ]$_m$
A
R2 may be fused
R3
ribose or ribose mimetic

prefered aromatic heterocycle, may be opened

| Collection of novel scaffold proposals from chemistry community | > 100 Library proposals | Ranking by
• Mapping onto Pharmacophore
• Chemical feasibility
• Patentability | 10 scaffolds selected for synthesis | Decoration of scaffolds using privileged fragments for purinergic GPCRs | 2,400 Cmpds synthesized |

b)

Chiral

c)

P2Y$_{12}$
Adenosin A1

Amine (6x)
Peptide (11x)
Lipid (3x)
Receptors

LIB1

LIB2

LIB3

LIB4

LIB5

Library compounds

Fig. 7.5 (a) Design process for proprietary scaffold libraries targeting puringeric GPCRs [34]. (b) Example of privileged fragments identified from known P2Y ligands and applied as building blocks for purinergic scaffold libraries. (c) Profiling of library sets targeting the P1 gene family [34]: the receptors are listed along the *x*-axis, and include the adenosine A1 receptor, the P2Y$_{12}$ receptor, six biogenic amine receptors, 11 peptide-binding GPCRs and three lipid-binding GPCRs. The library compounds are shown along the *y*-axis. Highlighted in yellow and orange are those compound showing more than 50 or 80% inhibition of radioligand binding at a compound concentration of 10 µM, respectively.

sign principles" for purinergic libraries (Fig. 7.5 a, top right, example of P1 agonists). These "design principles" resemble a construction plan to guide the invention of novel scaffolds. The example shown illustrates the molecular requirements of a P1 agonist scaffold: (i) it reveals the positions of the required hydrogen bond donor and hydrogen bond acceptor groups, (ii) provides tolerated options for the attachment of side chains and (iii) gives guidance on the number of atoms tolerated in the bicyclic ring system. Following these constraints chemists provided scaffold proposals meeting these molecular requirements of P1/P2Y (ant)agonists. These were ranked by chemical feasibility, by novelty, and by the quality of their mapping on the 3D pharmacophore leading to 10 scaffolds which were selected for synthesis. Fragments found to be privileged for the subfamily of purinergic GPCRs were used for "decoration" of the selected scaffolds: ligands of the purinergic GPCR subfamily were extracted from the MDDR and Aureus ligand databases and were subjected to a computational retrosynthetic analysis using the RECAP algorithm [13]. The ligands were dissected into their fragments, which were sorted by the frequency of their occurrence. Fragments most commonly found among P2Y antagonists and P1 agonists were used as building blocks for library synthesis (examples of P2Y privileged fragments are shown in Fig. 7.5 b). In total 2400 compounds were synthesized, all of which represent proprietary and lead-like compounds and thus offer excellent lead-finding opportunities for receptors of the family of purinergic GPCRs.

The libraries designed to target the subfamily of P1 receptors were evaluated by extensive profiling against a set of 22 GPCRs using radioligand displacement assays [34]. Of each of the five libraries diverse subsets of 20 to 25 compounds were selected for profiling. The result of the profiling is shown in Fig. 7.5 c and supports the design principles applied to P1 antagonists: (i) binding is mainly seen for the adenosine A1 receptor (some compounds show complete inhibition at a concentration of 10 μM), the only receptor of the P1 family among this receptor set; (ii) the affinity observed for other GPCR receptors is limited resulting in excellent selectivity profiles for most of the compounds with A1 affinity; (iii) three novel A1 antagonist scaffolds could thus be identified by profiling subsets taken from five scaffold libraries.

7.3
Understanding Molecular Recognition: Impact on GPCR Ligand Design

The example given above has illustrated how receptors with structurally-related physiological ligands can be grouped into subfamilies and how directed libraries (based on common pharmacophores and privileged building blocks) can be exploited for all members of the respective group of the gene family. This library design concept is applicable to those cases where the recognition site for the natural ligand is structurally conserved among the members of the respective subfamily. This is, however, often not the case, not even for receptors sharing the same physiological ligand. The prostaglandin DP and the chemoattractant

T-helper cell (CRTH2) receptors for example, are both activated by prostaglandin-D2, but do not share a similar conserved binding site for their natural ligand (nor do they share a significant overall sequence homology).

There is more and more evidence that the activity of most GPCRs can be modulated by ligands binding to a receptor site, similar to that used by retinal in rhodopsin [1]. The site is located within the 7TM domain and is used by many natural ligands such as the biogenic amines, for receptor binding and activation. In addition, it has been shown that many small molecule surrogate ligands of peptide- and chemokine-binding GPCRs appear to use this binding site to modulate the activity of their particular receptors. The existence of structurally related (sub)pockets conserved among several GPCRs is also reflected in the occurrence of common chemical structures found in GPCR ligands (see Section 7.2.3). In the context of library design and screening set compilation, it thus appears appropriate to compare and classify GPCRs based on the similarity of their putative ligand-binding site or subpockets, as this provides the best indication that two receptors would recognize similar ligands or similar ligand fragments. In order to support this classification method thorough analysis and recognition of the ligand–receptor interactions is therefore of utmost importance. Although there is still insufficient reliable 3D structural information on GPCRs and their ligand complexes, homology modeling combined with molecular recognition studies (mainly site-directed mutagenesis and affinity labeling) has provided valuable insights into GPCR structure and ligand–receptor interactions within recent years [37]. Typical receptor subsites for ligand binding or fingerprints on the primary sequence level for binding of the natural and surrogate ligands have been identified. The current approaches to and methods of understanding ligand recognition for the GPCR target family are presented in the following sections. In addition, examples are given of how these insights can be exploited for the design and synthesis of GPCR-directed libraries and screening sets.

7.3.1
Sites for Ligand Recognition within Biogenic-amine-binding and Other GPCRs

The first work describing ligand-binding subsites for monoamine-related GPCRs and correlating these with fragments commonly found within surrogate ligands was published by Jacoby and coworkers in 1999 [11, 38]. Based on site-directed mutagenesis data putative binding sites for common fragments of GPCR ligands were described. According to the three-binding-site hypothesis presented, biogenic amine receptors offer three binding (sub)sites to their typical ligands located within the 7TM domain. These sites are identified according to the name of their representative ligands: 5-hydroxytryptamine, propranolol, and 8-hydroxy-N,N-dipropylaminotetralin. According to Jacoby, these subsites overlap at an aspartate residue in helix TM3 (Asp3.32 using the Weinstein–Ballesteros nomenclature) [39], which constitutes the key anchor site for basic ligands of the biogenic amine receptors (Fig. 7.6). The existence of the three distinct binding sites is also reflected by the architectures of known high-affinity ligands which crosslink two or three "one-site-filling" fragments around a basic amino group.
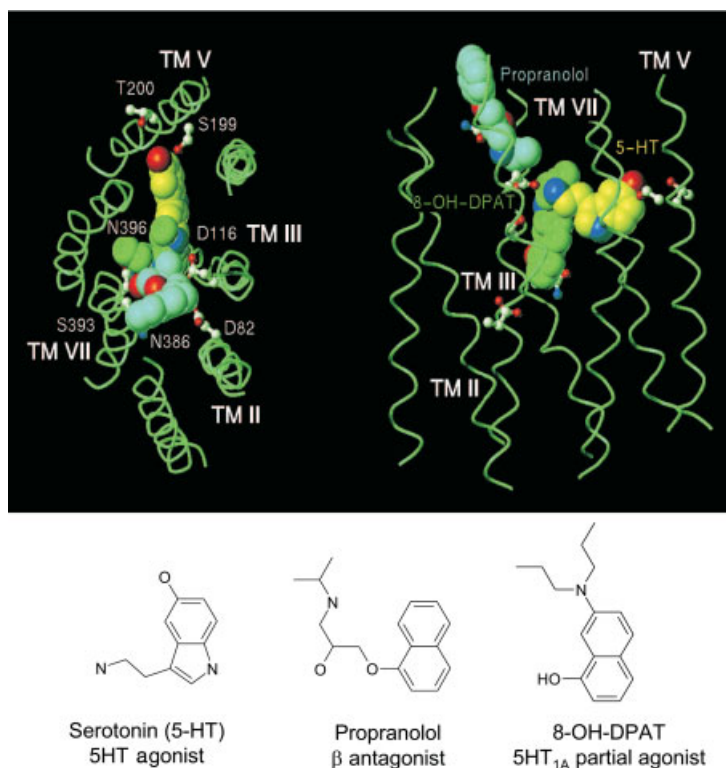
Serotonin (5-HT)
5HT agonist

Propranolol
β antagonist

8-OH-DPAT
5HT$_{1A}$ partial agonist

Fig. 7.6 Three-ligand binding sites of the 5-HT$_{1A}$ receptor in a rhodopsin-based 3D model according to Jacoby and co-workers (left, extracellular view; right, side view with extracellular side at the top) [38]. The three ligands are serotonin or 5-HT (yellow), propranolol (cyan) and 8-OH-DPAT (green). Residues identified by mutagenesis data are indicated. The „5-HT" site is located between TM3 (with Asp-116 as key recognition site) and TM5 (providing Ser-199 and Thr-200 to interact with the 5-OH group of serotonin).

A second site, the „propranolol" binding site is located between TM3 and TM7 (contributing e. g. Asn-386 to hydrogen bond the oxygen atoms of the oxy-propanolamine fragment in β-blockers). The third binding site, the „8-OH-DPAT" binding site is also located between TM3 and TM7. Ligands addressing this site, such as 8-OH-DPAT, are thought to be oriented parallel to the helices (interactions via 8-OH with Ser-393 and Asn-396 and via the amino group with Asp-116).

Knowledge of these ligand-recognition sites can serve two needs. First conserved ligand-binding subsites (versus overall sequence homology) can be used for predicting binding-site similarities and can thus support the identification of ligands for orphan receptors. Interestingly this novel "view" of receptor similarity based on the comparison of receptor sequences forming these subsites reveals the similarity of some peptide-binding GPCRs (such as the somatostatin receptor) and the biogenic amine GPCRs. This "view" is supported by the fact that these specific peptide-binding GPCRs indeed bind ligands which contain some struc-

tural fragments also found in typical biogenic amine-binding GPCR ligands. Second, receptors can be grouped according to conserved or similar binding sites in a more meaningful manner. Following the chemogenomics paradigm, known ligands of such a receptor group with predicted binding-site similarities can be identified and used as probes for the compilation of screening sets targeting novel receptors within this group. Also the ligand information from the defined GPCR subfamily can be used to design and synthesize libraries tailor-made for GPCRs with conserved or related binding (sub)sites.

7.3.2
Design of GPCR-directed Libraries Using "Motifs" and "Themes"

Whereas Jacoby and co-workers defined a GPCR similarity matrix by comparison of amino acids lining the putative (three) ligand binding sites, other approaches endeavor to detect fingerprints within GPCR sequences linked to the binding of the physiological or surrogate ligand. For some GPCRs there are conserved consensus signatures representing recognition sites for the natural ligand. For example a sequence motif characteristic of aminergic G protein-coupled receptors (composed of a conserved aspartic acid in TM3 and a conserved tryptophan residue in TM7) has recently been described [40]. The concept of recognition of consensus signatures within GPCR sequences and correlation with commonly found GPCR ligand fragments is applied by BioFocus [41]. Crossley and colleagues define "microenvironments" within GPCRs forming a set of amino acids that are commonly involved in ligand binding based on experimental molecular recognition data. From the identification of 30 to 40 critical residues, up to 14 of these microenvironments (called "themes" by the authors) have been identified. This work led to a re-classification of the class A GPCRs based on the recognition of a receptor-specific fingerprint showing the presence or absence of these themes. Using multidimensional scaling on this high-dimensional fingerprint matrix the authors arrive at a two-dimensional representation of new GPCR clusters. On the ligand side GPCR privileged fragments, so-called "motifs", recognized by these "themes" have been identified from the wealth of known GPCR ligands. Closeness, i.e. similarity, in the receptor projection suggests that the receptors share similar ligand recognition sites and are likely to interact with similar chemical motifs. The use of this new approach to guide the selection of building blocks and to generate libraries tailor-made for the targeted receptors has been reported and the first results of this attractive chemogenomics approach to lead finding are eagerly awaited.

7.3.3
"Chemoprints" for Recognition of GPCR-privileged Fragments

Only recently, researchers from Novo Nordisk provided structural views of the molecular recognition between several GPCR-privileged fragments and the respective receptor binding sites. Docking studies of privileged structures containing GPCR ligands into rhodopsin-derived homology models provided the putative binding
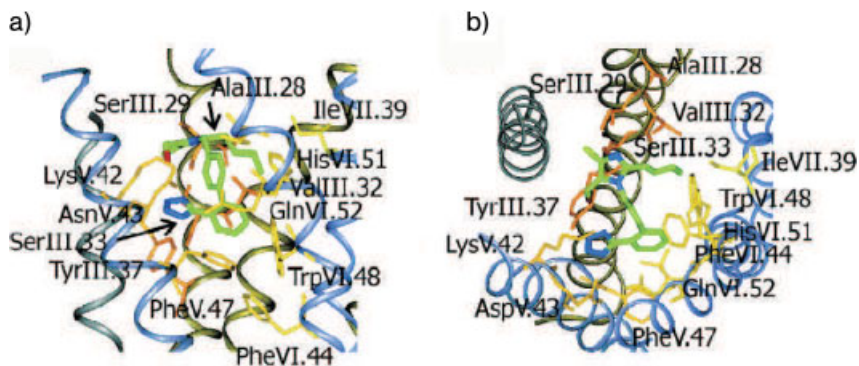
Fig. 7.7 Putative ligand-binding pocket of the type-1 angiotensin II (AT1) receptor complexed with losartan [42]. (A) Side view with TM5, TM6, and TM7 in front, and (B) top view from the extracellular side. Ligand atoms are color-coded according to atom types: carbon, green; oxygen, red; nitrogen, blue; and chloride, light yellow. Receptor atoms are shown in yellow and TMs are shown in shades of brown (TM1) to blue (TM7).
(Reprinted with permission from [42]; copyright 2004 American Chemical Society).

Chemoprint	3.32	5.39	5.42	5.43	5.46	5.47	6.48	6.51	6.52	6.55
	Asp					Phe	Trp	Phe		
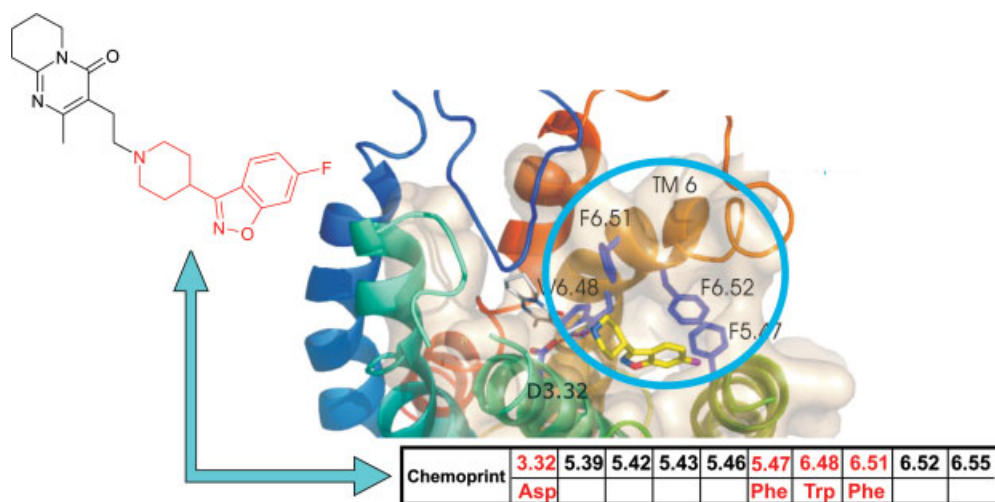
Fig. 7.8 Putative binding mode of risperidone in the α_{1a} adrenergic receptor [43]. The model is in line with site-directed mutagenesis data and suggests that the "chemoprint" for recognition of 4-aryl piperidines and piperazines is formed by residues Asp3.32, Phe5.47, Trp6.48 and Phe6.51.

modes for phenyl-indole-, spiro-piperidine-indane-, and 2-tetrazole-biphenyl-containing ligands (Fig. 7.7) [42]. The authors identified a conserved binding region spanned by two phenylalanines Phe5.47 and Phe6.44 and a tryptophan residue Trp6.48 harboring these GPCR-privileged fragments.

GPCR homology modeling, docking and experimental profiling can be used to identify motifs within GPCR sequences, so-called "chemoprints" that are linked to the recognition of GPCR-privileged fragments. These chemoprints allow the identification of relevant receptor similarities and support the design of compound libraries tailor-made for the target GPCR. Docking of compounds containing GPCR-privileged fragments into rhodopsin-derived homology models provide an initial hypothesis regarding the interaction of these receptor–ligand pairs. Figure 7.8 shows a homology model of the α_{1a} adrenergic receptor together with the putative binding mode of risperidone [43], a $5HT_{2A}$ antagonist with nanomolar affinity towards the α_{1a} adrenergic receptor. The binding mode is in line with site-directed mutagenesis data supporting the interaction of the ligand with the aspar-
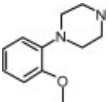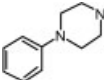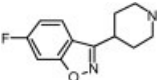
Table 7.1 Patchwork alignment of GPCR sequences (biogenic amine and peptide-binding GPCRs). The alignment reveals the presence or absence of the „chemoprint" composed of Asp3.32, Phe/Tyr5.47, Trp6.48 and Phe/Tyr6.51 (in bold), which is linked to the recognition of 4-aryl piperidines and piperazines. Residue numbers are given according to the Weinstein nomenclature.

	3.32	5.39	5.42	5.43	5.46	5.47	6.48	6.51	6.52	6.55	Chemoprint present
Adrenergic α_{1A}	**Asp**	Val	Ser	Ala	Ser	**Phe**	**Trp**	**Phe**	Phe	Met	Yes
Dopaminergic D2	**Asp**	Val	Ser	Ser	Ser	**Phe**	**Trp**	**Phe**	Phe	His	Yes
Histaminergic H1	**Asp**	Lys	Thr	Ala	Asn	**Phe**	**Trp**	**Tyr**	Phe	Phe	Yes
Histaminergic H3	**Asp**	Leu	Ala	Ser	Glu	**Phe**	**Trp**	**Tyr**	Thr	Met	Yes
Muscarinic M1	**Asp**	Thr	Thr	Ala	Ala	**Phe**	**Trp**	**Tyr**	Asn	Val	Yes
Muscarinic M3	**Asp**	Thr	Thr	Ala	Ala	**Phe**	**Trp**	**Tyr**	Asn	Val	Yes
Serotonin $5HT_{2C}$	**Asp**	Val	Gly	Ser	Ala	**Phe**	**Trp**	**Phe**	Phe	Asn	Yes
Serotonin $5HT_{2A}$	**Asp**	Val	Gly	Ser	Ser	**Phe**	**Trp**	**Phe**	Phe	Asn	Yes
Endothelin A	Glu	Met	Tyr	Phe	Tyr	**Phe**	**Trp**	Glu	Asn	Asn	No
Galanin 1	Phe	Val	Thr	Phe	Gly	**Phe**	**Trp**	His	His	His	No
Angiotensin 1	Leu	Ala	Lys	Asn	Gly	**Tyr**	**Trp**	**Phe**	His	Thr	No
Bradykinin 2	Ile	Asn	Leu	Asn	Gly	**Phe**	**Trp**	**Phe**	Gln	Thr	No
Melanocortin 3	Ile	Val	Cys	Leu	Met	**Phe**	**Trp**	**Phe**	Phe	Leu	No
Neurokinin 1	Pro	His	Val	Thr	Ile	**Tyr**	**Trp**	**Phe**	His	Phe	No
Chemokine CCR1	Tyr	Ala	Leu	Asn	Gly	Leu	**Trp**	**Tyr**	Asn	Ile	No
Choleocystokinin B	Met	Ser	Leu	Leu	Leu	**Phe**	**Trp**	Val	Tyr	Asn	No
Opiate μ	**Asp**	Lys	Val	Phe	Ala	**Phe**	**Trp**	Ile	His	Val	Modified
Neuropeptide Y	Gln	Thr	Leu	Leu	Gln	**Tyr**	**Trp**	Leu	Thr	Asn	No
Vasopressin 1a	Gln	Gly	Phe	Val	Val	Val	**Trp**	**Phe**	Phe	Gln	No
MCH 1	**Asp**	Thr	Gln	Phe	Ala	**Phe**	**Trp**	**Tyr**	Tyr	Gln	Yes
Somatostatin Sst2	**Asp**	Ile	Thr	Phe	Gly	**Phe**	**Trp**	**Phe**	Tyr	Asn	Yes
Urotensin II	**Asp**	Leu	Leu	Phe	Ser	Ile	**Phe**	**Phe**	Trp	Gln	Modified

tate residue in TM3 (Asp3.32) and the positioning of the 4-aryl-piperidine moiety into the subpocket formed by hydrophobic residues of helices TM5 and TM6. From this docking mode the hypothesis can be derived that 4-aryl piperidines and piperazines, commonly found fragments within GPCR ligands (see Fig. 7.3), can bind to such GPCRs, which provide the *chemoprint* composed of the residue set Asp3.32, Phe5.47, Trp6.48 and Phe6.51 as the recognition site.

A sequence comparison of several biogenic amine receptors (see Table 7.1) shows that it is not only the aspartate residue in position Asp3.32 and the tryptophan residue in Trp6.48, both of which were found in the consensus signature for biogenic amine receptors by Pfizer [40], that are conserved. In addition, the two aromatic residues in positions Phe5.47 and Phe/Tyr6.51 are present in all biogenic amine receptor sequences shown in Table 7.1. The presence of this *chemoprint* suggests that 4-aryl piperidines and piperazines are suited to be building blocks for libraries targeting biogenic amine receptors. To test this hypothesis we profiled 25 piperazines and piperidines commonly found within GPCR ligands against a set of eight biogenic amine receptors [34]. Table 7.2 shows the results for

Table 7.2 Binding affinity profile of GPCR-privileged fragments tested against a set of biogenic amine-binding and peptide-binding GPCRs [34]. The percent binding values of three fragments tested against a set of eight biogenic amine- and 11 peptide-binding GPCRs using radioligand displacement assays (compound concentration of 10 µM) are shown. Fragment–receptor pairs with significant affinity (> 40 % of binding) are highlighted in bold.

Adrenergic α_{1A}	**97**	**48**	**97**
Dopaminergic D2	**87**	3	**61**
Histaminergic H1	**76**	9	**61**
Histaminergic H3	18	6	19
Muscarinic M1	**65**	25	**45**
Muscarinic M3	**73**	12	39
Serotonin $5HT_{2C}$	**91**	**45**	**73**
Serotonin $5HT_{2A}$	**97**	**68**	**100**
Endothelin A	6	12	7
Galanin 1	0	0	–4
Angiotensin II, type 1	6	2	5
Bradykinin B2	–5	2	12
Melanocortin MC3	4	11	12
Neurokinin NK1	22	–2	5
Chemokine CXCR2	–3	–3	–12
Choleocystokinin CCK B	17	7	13
Opiate	**54**	23	9
Neuropeptide NPY	4	2	–3
Vasopressin V1a	11	18	16

three GPCR-privileged fragments. The good affinity of these fragments found for the biogenic amine GPCRs indicates that they are suitable building blocks for libraries targeting this class of GPCRs.

In order to further validate the correlation between the occurrence of the 4-aryl piperidine/piperazine-recognizing chemoprint and experimental affinity we profiled the fragments against a set of 11 peptide-binding GPCRs [34]. The results for three fragments are shown in Table 7.2. Indeed, some fragments were found to reveal moderate affinity towards the opiate receptor (greater than 50% inhibition; binding affinities in the range of $K_i = 2$–$10\,\mu M$). The patchwork alignment in Table 7.1 illustrates that the opiate receptor is the only receptor among the 11 receptors within the panel, that has an aspartate residue in position 3.32 which provides a hydrophobic patch of residues from helices in TM5 and TM6 (with Ile in 6.51 instead of Phe). Interestingly, no significant affinity was seen for any of the other 10 peptide-binding GPCRs lacking the corresponding chemoprint. This result confirms the chemoprint hypothesis linking a sequence motif to a privileged chemical fragment. It also suggests that 4-aryl piperidines/piperazines and their spiro derivatives not only represent excellent building blocks for the design of libraries targeting the family of biogenic amine receptors and the opiate receptor, but that they are also suitable for libraries targeting other peptide-binding GPCRs sharing the respective chemoprint, such as the somatostatin receptors, the MCH receptor and the UII receptor (see Table 7.1). This example illustrates how experimentally validated chemoprints, linking sequence motifs with GPCR-privileged structures, can be used to guide the selection of building blocks and scaffolds in the design of libraries targeting novel GPCRs.

7.3.4
Molecular Interaction Models by Proteochemometrics

Another computational approach to mapping molecular interaction space without knowledge of the 3D structure and known as proteochemometrics, has been followed by the Wikberg group [5, 44]; this is an extension of traditional ligand-based 3D QSAR approaches. It exploits affinity data for series of diverse organic ligands binding to different receptor subtypes, correlating it to descriptors and cross-terms derived from amino acid sequences (e.g. z scales of amino acids located within the putative binding site) and the chemical structures of their ligands. Wikberg and his team have performed several studies on GPCRs, which used profiling data of compound series against a set of receptors (or mutants of the same receptor) and subsequent proteochemometric analysis to provide insights into ligand recognition. Thus, ligand–receptor interaction models have been obtained for the subfamily of melanocortin receptors (rationalizing the binding of peptidic ligands) [45], for the biogenic amine receptors (based on profiling data from 23 ligands on 21 receptors) [12] and finally for the alpha adrenergic receptor (using the SAR of a ligand series profiled on a set of mutant receptor forms) [46]. Statistically valid models resulted in all cases and these could be used to make sound experimental predictions while evaluation of these models

gave important insight into the interaction mode of GPCRs with their ligands. A word of caution might be necessary here as only those receptor interaction sites for which variations are present within the training set can be identified within the resulting cross-target models, in a manner similar to that seen in classical structure–activity studies where derivatives of one chemical series are tested against one receptor. Those residues within the training set of (mutant) receptors, which have not been varied and experimentally tested will not be identified by these models.

7.4
Outlook

Although powerful bio- and chemoinformatic tools have been developed to provide the desired link between chemical and biological spaces, the success of exploiting this information for the design of GPCR-targeted libraries will clearly depend on the availability of further valid structural information on GPCRs. The elucidation of 3D structures of several membrane proteins within recent years (e.g. cytochrome C oxidase, bacteriorhodopsin, potassium channel, bovine rhodopsin, photosynthetic reaction center) and the enormous effort currently being invested in the structural genomics of GPCRs by industrial and academic groups supports some realistic optimism that this challenging task will be fulfilled in the near future [47]. *Chemoprints* could then be directly derived from experimental 3D structures of GPCRs complexed with ligands containing GPCR-privileged motifs and libraries could be tailor-made for a novel GPCR. These targeted-libraries using building blocks and scaffolds tuned for the specific receptor could support lead finding especially in cases where traditional high-throughput screening failed to deliver chemical starting points for drug discovery.

Acknowledgments

References

1 Klabunde, T., Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **2002**; *3*: 928–944.

2 Venter, J. C. et al. The sequence of the human genome. *Science* **2001**; *291*: 1304–1351.

3 Wise, A., Gearing, K., Rees, S. Target validation of G-protein coupled receptors. *Drug Discovery Today* **2002**; *7*: 235–246.

4 Caron, P. R., Mullican, M. D., Mashal, R. D., Wilson, K. P., Su, M. S., Murcko, M. A. Chemogenomic approaches to drug discovery. *Curr Opin Chem Biol* **2001**; *5*: 464–470.

5 Kubinyi, H., Müller, G. (Eds). *Chemogenomics in Drug Discovery*, Wiley-VCH: Weinheim, **2004**.

6 Bredel, M., Jacoby, E. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat Rev Genet* **2004**; *5*: 262–275.

7 Jimonet, P., Jäger, R. Strategies for designing GPCR-focused libraries and screening sets. *Curr Opin Drug Discovery Dev* **2004**; *7*: 325–333.

8 Wess, G., Urmann, M., Sickenberger, B. Medicinal chemistry: challenges and opportunities. *Angew Chem Int Ed Engl* **2001**; *40*: 3341–3350.

9 Wermuth, C. G. Selective optimization of side activities: another way for drug discovery. *J Med Chem* **2004**; *47*: 1303–1314.

10 Krejsa, C. M., Horvath, D., Rogalski, S. L., Penzotti, J. E., Mao, B., Barbosa, F., Migeon, J. C. Predicting ADME properties and side effects: the BioPrint approach. *Curr Opin Drug Discovery Dev* **2003**; *6*: 470–480.

11 Jacoby, E., Schuffenhauer, A., Floersheim, P. Chemogenomics knowledge-based strategies in drug discovery. *Drug News Perspect* **2003**; *16*: 93–102.

12 Lapinsh, M., Prusis, P., Lundstedt, T., Wikberg, J. E. Proteochemometrics modeling of the interaction of amine G-protein coupled receptors with a diverse set of ligands. *Mol Pharmacol* **2002**; *61*: 1465–1475.

13 Lewell, X. Q., Judd, D. B., Watson, S. P., Hann, M. M. M. RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* **1998**; *38*: 511–522.

14 Schuffenhauer, A., Floersheim, P., Acklin, P., Jacoby, E. Similarity metrics for ligands reflecting the similarity of the target proteins. *J Chem Inf Comput Sci* **2003**; *43*: 391–405.

15 Pearlman, R. S., Smith, K. M. Metric Validation and the receptor-relevant subspace concept. *J Chem Inf Comput Sci* **1999**; *39*: 28–35.

16 Rarey, M., Dixon, J. S. Feature trees: a new molecular similarity measure based on tree matching. *J Comput Aided Mol Des* **1998**; *12*: 471–490.

17 Schneider, G., Neidhart, W., Giller, T., Schmid, G. "Scaffold-hopping" by topological pharmacophore search: a contribution to virtual screening. *Angew Chem Int Ed Engl* **1999**; *38*: 2894–2896.

18 Schneider, G., Schneider, P. Navigation in chemical space: ligand-based design of focused compound libraries. In *Chemogenomics in Drug Discovery*, Kubinyi, H., Müller, G. (Eds), Wiley-VCH, Weinheim, **2004**.

19 Schmitt, S., Kuhn, D., Klebe, G. A new method to detect related function among proteins independent of sequence and fold homology. *J Mol Biol* **2002**; *323*: 387–406.

20 Schuffenhauer, A., Jacoby, E. Annotating and mining the ligand–target chemogenomics knowledge space. *Drug Discovery Today* **2004**; *2*: 190–200.

21 Balakin, K. V., Tkachenko, S. E., Lang, S. A., Okun, I., Ivashchenko, A. A., Savchuk, N. P. Property-based design of GPCR-targeted library. *J Chem Inf Comput Sci* **2002**; *42*: 1332–1342.

22 Mason, J. S., Morize, I., Menard, P. R., Cheney, D. L., Hulme, C., Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of

the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J Med Chem* **1999**; *42*: 3251–3264.

23 Lamb, M. L., Bradley, E. K., Beaton, G., Bondy, S. S., Castellino, A. J., Gibbons, P. A., Suto, M. J., Grootenhuis, P. D. J. Design of a gene family screening library targeting G-protein coupled receptors. *J Mol Graphics Model* **2004**; *23*: 15–21.

24 Manallack, D. T., Pitt, W. R., Gancia, E., Montana, J. G., Livingstone, D. J., Ford, M. G., Whitley, D. C. Selecting screening candidates for kinase and G protein-coupled receptor targets using neural networks. *J Chem Inf Comput Sci* **2002**; *42*: 1256–1262.

25 Lavrador, K., Murphy, B., Saunders, J., Struthers, S., Wang, X., Williams, J. A screening library for peptide activated G-protein coupled receptors. 1. The test set. *J Med Chem* **2004**; *47*: 6864–6874.

26 Evans, B. E., Rittle, K. E., Bock, M. G., DiPardo, R. M., Freidinger, R. M., Whitter, W. L., Lundell, G. F., Veber, D. F., Anderson, P. S., Chang, R. S. Methods for drug discovery: development of potent, selective, orally effective cholecystokinin antagonists. *J Med Chem* **1988**; *31*: 2235–2246.

27 Gordeev, M. F., Patel, D. V., England, B. P., Jonnalagadda, S., Combs, J. D., Gordon, E. M. Combinatorial synthesis and screening of a chemical library of 1,4-dihydropyridine calcium channel blockers. *Bioorg Med Chem* **1998**; *6*: 883–889.

28 Neustadt, B. R., Smith, E. M., Lindo, N., Nechuta, T., Bronnenkant, A., Wu, A., Armstrong, L., Kumar, C. Construction of a family of biphenyl combinatorial libraries: structure–activity studies utilizing libraries of mixtures. *Bioorg Med Chem Lett* **1998**; *8*: 2395–2398.

29 Thompson, L. A., Ellman, J. A. Synthesis and applications of small molecule libraries. *Chem Rev* **1996**; *96*: 555–600.

30 Horton, D. A., Bourne, G. T., Smythe, M. L. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem Rev* **2003**; *103*: 893–930.

31 Guo, T., Hobbs, D. W. Privileged structure-based combinatorial libraries targeting G protein-coupled receptors. *Assay Drug Dev Technol* **2003**; *1*: 579–592.

32 Berk, S. C., Rohrer, S. P., Degrado, S. J., Birzin, E. T., Mosley, R. T., Hutchins, S. M., Pasternak, A., Schaeffer, J. M., Underwood, D. J., Chapman, K. T. A combinatorial approach toward the discovery of non-peptide, subtype-selective somatostatin receptor ligands. *J Comb Chem* **1999**; *1*: 388–396.

33 Rohrer, S. P., Birzin, E. T., Mosley, R. T., Berk, S. C., Hutchins, S. M., Shen, D. M., Xiong, Y., Hayes, E. C., Parmar, R. M., Foor, F., Mitra, S. W., Degrado, S. J., Shu, M., Klopp, J. M., Cai, S. J., Blake, A., Chan, W. W., Pasternak, A., Yang, L., Patchett, A. A., Smith, R. G., Chapman, K. T., Schaeffer, J. M. Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry. *Science* **1998**; *282*: 737–740.

34 Klabunde, T., Jäger, R. Design of GPCR targeted compound collections. *Proc Erst Schering Research Foundation* (in press).

35 MDL Information Systems Inc. www.mdli.com.

36 Aureus Pharma. www.aureus-pharma.com.

37 Shi, L., Javitch, J. A. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu Rev Pharmacol Toxicol* **2002**; *42*: 437–467.

38 Jacoby, E., Fauchère, J.-L., Raimbaud, E., Ollivier, S., Michel, A., Spedding, M. A three binding site hypothesis for the interaction of ligands with monoamine G protein-coupled receptors: implications for combinatorial ligand design. *Quant Struct-Act Relat* **1999**; *18*: 561–571.

39 Ballosteros, J. A., Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations of G protein-coupled receptors. *Methods Neurosci* **1995**; *25*: 366–428.

40 HUANG, E. S. Construction of a sequence motif characteristic of aminergic G protein-coupled receptors. *Prot Sci* **2004**; *12*: 1360–1367.

41 CROSSLEY, R., ROSE, V., STEVENS, A. P. Compound libraries. Patent WO03004147, **2003**.

42 BONDENSGAARD, K., ANKERSEN, M., THOGERSEN, H., HANSEN, B. S., WULFF, B. S., BYWATER, R. P. Recognition of privileged structures by G-protein coupled receptors. *J Med Chem* **2004**; *47*: 888–899.

43 EVERS, A., KLABUNDE, T. Structure-based drug discovery using GPCR homology modelling: successful virtual screening for antagonists of the alpha1 a receptor. *J Med Chem* **2005**; *48*: 1088–1097.

44 WIKBERG, J. E. S., LAPINSH, M., PRUSIS, P. Proteochemometrics: a tool for modeling the molecular interaction space. In *Chemogenomics in Drug Discovery*; Kubinyi, H.; Müller, G. (Eds), Wiley-VCH: Weinheim, **2004**.

45 PRUSIS, P., MUCENIECE, R., ANDERSSON, P., POST, C., LUNDSTEDT, T., WIKBERG, J. E. PLS modeling of chimeric MS04/MSH-peptide and MC1/MC3-receptor interactions reveals a novel method for the analysis of ligand–receptor interactions. *Biochim Biophys Acta* **2001**; *1544*: 350–357.

46 LAPINSH, M., PRUSIS, P., GUTCAITS, A., LUNDSTEDT, T., WIKBERG, J. E. Development of proteo-chemometrics: a novel technology for the analysis of drug–receptor interactions. *Biochim Biophys Acta* **2001**; *1525*: 180–190.

47 LUNDSTROM, K. Structural genomics on GPCRs: the MePNet approach. *Curr Drug Discovery* **2002**; *5*: 29–33.

8
Strategies for the Design of pGPCR-targeted Libraries

Nikolay P. Savchuk, Sergey E. Tkachenko, and Konstantin V. Balakin

8.1
Introduction

Peptidergic G protein-coupled receptors (pGPCRs) are a diverse group within the GPCR superfamily, their ligands being endogenous peptides. Binding of these peptides to the extracellular or transmembrane regions causes conformational changes of the receptor that act as a switch transferring the signal to the trimeric guanine nucleotide binding regulatory proteins (G proteins), thus inhibiting or stimulating the production of intracellular secondary messengers, such as cyclic adenosine monophosphate (cAMP) or Ca^{2+} ions. As the members of the super-family of G protein-coupled receptors which represent the most successful drug targets [1], peptidergic GPCRs possess excellent therapeutic potential. It is be-lieved that therapeutic intervention at these receptors will have major benefits in a wide range of human diseases.

Information concerning these receptors and their endogenous ligands is highly important for the design of novel drugs for this target family. In the first part of this chapter, the nomenclature and function of pGPCRs, their endogenous ligands and therapeutic potential are surveyed in brief.

8.1.1
Peptidergic GPCRs: Brief Overview

Endogenous peptides exert their biological functions through interaction with specific high- and low-affinity pGPCRs. At least 35 different families of pGPCRs and their ligands have been identified to date, with multiple receptor subtypes in most of these families each of which are encoded by separate genes [2]. The major-ity of families can be grouped by sequence similarity within the broad classifica-tion scheme for GPCRs as Class A (rhodopsin like), whereas the rest can be placed into Class B (secretin/glucagon receptor-like; Table 8.1). Class A receptors are dis-tinguished by a set of highly conserved amino acids in the cytoplasmic half of the

Table 8.1 Major pharmacologically significant peptidergic GPCR families and their therapeutic potential.

Families	Receptors	Required ligands	Therapeutic areas
Class A			
• Angiotensin	AT1, AT2	Antagonists	Hypertension, pain, glaucoma, cognition disorders,
• Bombesin	BRS-3, GRPR, NMBR	Antagonists	Cancer, renal diseases
• Bradykinin	B1R, B2R	Antagonists	Pain, allergy, asthma, cancer
• Chemokine	CCR1 – CCR5, CCR9, CXCR1 – CXCR5	Antagonists	Arthritis, cancer, Alzheimer's disease, atherosclerosis, asthma, pulmonary disorders, allergy, human immuno-deficiency virus infection
• Cholecystokinin	CCK-AR, CCK-BR	Agonists/ antagonists	Pancreatic disorders, obesity, diabetes, CNS disorders, pain
• Endothelin	ETA-R, ETB-R	Agonists/ antagonists	Hypertension, cardiovascular disorders, cancer
• Galanin	GAL1-R – GAL3-R	Agonists/ antagonists	CNS disorders, pain, obesity, cognition disorders
• Melanin-concentrating hormone	MCH-R1, MCH-R2	Antagonists	Obesity, diabetes, CNS disorders
• Melanocortin	MC1-R – MC5-R	Agonists/ antagonists	eating disorders, obesity, cachexia, anorexia, CNS disorders
• Neuropeptide Y	NPY1-R – NPY5-R	Agonists/ antagonists	Obesity, hypertension, stroke, diabetes
• Neurotensin	NTSR1, NTSR2	Agonists/ antagonists	CNS disorders, Parkinson's disease, gastrointestinal disorders, pain
• Opioid	DOR-1, KOR-1, MOR-1	Agonists/ antagonists	Pain, CNS disorders, cardiovascular disorders
• Opioid receptor-like	ORL1	Agonists/ antagonists	Pain, CNS disorders, cardiovascular disorders
• Orexin	OX1R, OX2R	antagonists	Obesity, sleep disorders
• Oxytocin	OTR	Agonists/ antagonists	Sexual and gynecological disorders
• Somatostatin	SSTR1 – SSTR5	Agonists/ antagonists	Diabetes, cancer, cardiovascular disorders, cognition disorders, diarrhea, glaucoma, pancreatic disorders, acromegaly

Table 8.1 (continued)

Families	Receptors	Required ligands	Therapeutic areas
Class A			
• Tachykinin	NK-1R – NK-3R	Antagonists	Allergy, asthma, migraine, CNS disorders, arthritis, pulmonary diseases, nausea and vomiting
• Urotensin II	UR-II-R	Antagonists	Angina pectoris, cardiovascular disorders
• Vasopressin	V1R,V1aR,V1bR, V2R	Agonists/ antagonists	Urinary incontinence, renal diseases, diabetes, hyper-tension, dysmenorrhea, CNS disorders, cardiovascular disorders
Class B			
• Calcitonin	CTR	Agonists	Osteoporosis, diabetes
• Glucagon	GLR	Antagonists	Diabetes, obesity
• Glucagon-like peptide 1	GLP-1-R	Agonists/ antagonists	Diabetes, obesity

seven-transmembrane core that are required for receptor stability and for mediating the conformational changes that underlie receptor activation [3, 4]. Despite similar topology, Class B receptors share little amino acid sequence similarity with Class A receptors and are distinguished by a large extracellular amino-terminal domain that is critical for ligand binding and which contains six highly conserved cysteine residues that are likely involved in disulfide bond formation [5].

Receptors within each family are presumed to have similar binding domains specific for their cognate class of signaling peptides and, in many cases, activate common intracellular signaling pathways. The specificity of G protein coupling to a given GPCR is typically defined in terms of the class of G_α subunit (G_s, $G_{i/o}$, $G_{q/11}$, G_{12}). Class A pGPCRs appear to couple primarily to either $G_{q/11}$ and/or $G_{i/o}$, which are commonly associated with activation of phospholipase C and inhibition of adenylyl cyclase, respectively. In contrast, Class B pGPCRs are primarily coupled to G_s and mediate stimulation of adenylyl cyclase. Many pGPCRs have been shown to couple to more than one G protein subtype [6].

Receptors within each family differ in their relative affinity for ligands within a peptide family, G protein-coupling specificity, desensitization kinetics, and ability to associate with other GPCRs [7–9].

8.1.2
Endogenous Ligands for pGPCRs

Endogenous peptide ligands for GPCRs are usually generated by enzymatic cleavage of a prepro-precursor to give fragments from four to 90 amino acids in length. Several biologically active peptides specific for one or more GPCRs within a family can be generated from a single precursor, such as somatostatin-14 and -28 from prepro-somatostatin [10, 11]. Additional mechanisms include processing of peptides or proteins that are already biologically active. For example, enzyme-mediated cleavage of the C-terminal dipeptide motif from AT1 produces AT2 [12].

Posttranslational modification of amino acid residues in peptide ligands can regulate their potency or specificity. For peptides such as CCK, gastrin, NPY, orexins A and B, and the pro-opioid melanocortin-derived peptides, C-terminal α-amidation is required for function [13, 14]. Octanoylation of ghrelin at Ser^3 is required for biological activity [15] and sulfation of Tyr^7 or Ala^7 in CCK and gastrin respectively is required for full potency of action on CCK1 receptors [14]. Also, neuropeptide B, a ligand recently identified for the orphan GPCR GPR7, was shown to be brominated specifically at an N-terminal Trp residue [16].

Different pGPCRs display specific patterns of ligand selectivity [9]. Members of several pGPCR families display specific preferences for certain ligands. Thus, the SSTR5 subtype of the somatostatin receptors is selective for somatostatin-28 [17], unlike the other four SSTR isoforms, which appear to preferentially bind somatostatin-14 [18]. In many cases, one receptor subtype can bind only one specific ligand, whereas another subtype displays equally high affinity for a number of endogenous ligands [19, 20]. It can be hypothesized that such a diversity of peptides available to activate a particular pGPCR could allow a greater dynamic range of receptor responses while maintaining levels of specificity for specific intercellular signals.

Recently, several interesting examples of functional differences between ligands that can act on the same receptor were reported. Thus, an endogenous truncated NKA peptide lacking the first three amino acids (NKA(4–10)) evokes only the calcium response and not the cAMP response through NK2 receptors [8]. Similar effects described for several other pGPCRs [21] suggest that different conformations of the receptor could underlie distinct activation states, which have different ligand affinities, resulting in coupling to different G proteins. Cooperative action of distinct ligands on the same receptor has also been reported [22].

8.1.3
Potential Therapeutic Targets of pGPCRs

Peptidergic GPCRs have been shown to be involved in many physiological functions. Many disease conditions have been related to the malfunction of pGPCRs, including cardiovascular, neurologic, gastrointestinal, autoimmune and ocular diseases, human immunodeficiency virus (HIV) infection, obesity and cancer (Table 8.1). Comprehensive reports of the potential pharmacological activities of
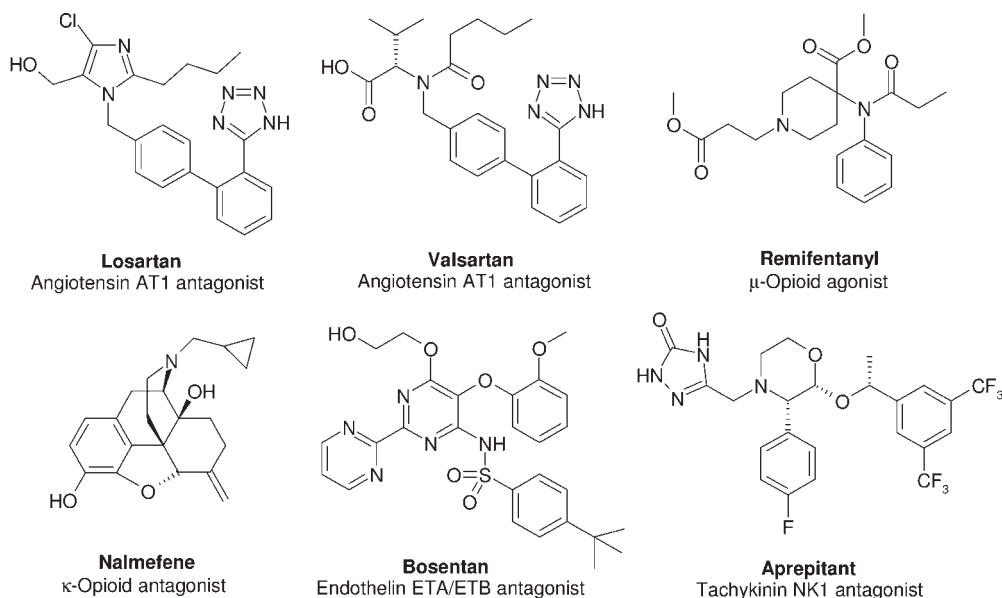
Fig. 8.1 Examples of pGPCR-active drugs which have been marketed.

the main pGPCR families have been reported elsewhere. This information provides a very strong rationale for the implementation of drug discovery programs directed at pGPCRs.

There are several drugs which have been launched onto the market in the past decade acting through modulation of pGPCR function (Fig. 8.1). Among these is a series of structurally related angiotensin antagonists, such as losartan and valsartan, opioid antagonists remifentanyl and nalmefene, endothelin antagonist bosentan and tachykinin antagonist aprepitant. Numerous examples of small-molecule pGPCR ligands (Fig. 8.2), which have progressed to clinical status, also validate the significant allocation of resources to this research area (the data are from the Prous Ensemble database (Prous Science, 2004; http://www.prous.com/)).

8.2
Approaches to the Design of pGPCR-targeted Libraries

High-throughput screening of large diversity-based libraries is still a common strategy within many pharmaceutical companies for the discovery of pGPCR leads. Several examples of successful pGPCR leads resulting from HTS campaigns can be found in the scientific literature. Recent examples include antagonists of chemokine receptors CCR3 [23] and CCR5 [24], gonadotropin-releasing hormone receptor [25], neurokinin-1 receptor [26], bradykinin [27] and neuropeptide Y receptors [28]. However, as noted by many researchers in the field, there is
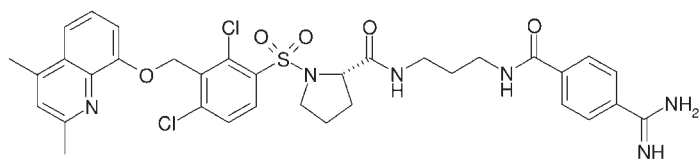
Anatibant (Phase I)
Bradykinin B2 antagonist
Fournier
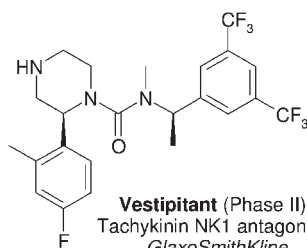
Vestipitant (Phase II)
Tachykinin NK1 antagonist
GlaxoSmithKline
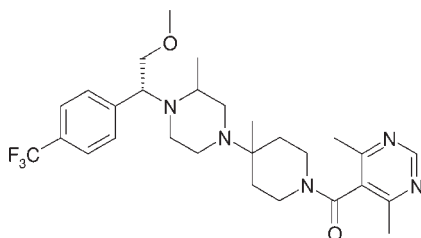
SCH-D (Phase I)
Chemokine CCR5 antagonist
Schering-Plough
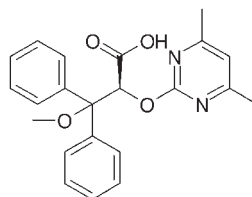
Ambrisentan (Phase III)
Endothelin ETA antagonist
Abbott
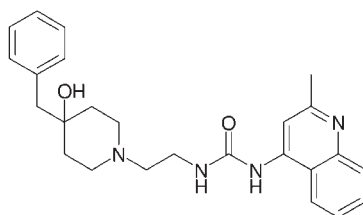
Palosuran (Phase II)
Urotensin-II antagonist
Actelion

Fig. 8.2 Examples of pGPCR-active agents entered into clinical trials.
The data shown reflects the status of these agents at the end of 2004.

no evidence that high-throughput technologies, including parallel synthesis/combinatorial chemistry and HTS provided the expected impedance to the lead discovery process [1, 29]. Therefore, a number of approaches have been used for the design of more focused screening libraries [30]. These range from pharmacophore and target structure-based design through combinatorial approaches to mimetics of natural ligands to methods based on intelligent data mining techniques. Chemogenomic approaches provide novel opportunities for the design of targeted libraries through better understanding of the relationships between pGPCR sequence and compounds that interact at particular receptors.

Here we highlight some of the strategies mentioned, which we have found to be useful for the design of pGPCR-targeted libraries. The examples presented demonstrate how the knowledge obtained from receptor–ligand interaction models and structures of known ligands, both endogenous and exogenous, can be applied to the design of small-molecule mimetics of peptide GPCR ligands. But before moving forward, we highlight some of the typical problems associated with the discovery of pGPCR active drugs.

8.2.1
Problems in Drug Discovery Directed Towards pGPCRs

Historically, the discovery of drugs acting on non-peptidergic GPCRs (non-pGPCRs) has been more successful as compared to pGPCRs. Research on angiotensin and opioid receptors yielded therapies that display notable therapeutic and market benefits. Progress in drug discovery against the remaining members of this family has been less successful. To date, only two synthetic ligands to other pGPCRs, namely bosentan (endothelin ETA/ETB antagonist) and aprepitant (tachykinin NK1 antagonist; Fig. 8.2), have reached the market. These data suggest that (1) the potential of pGPCRs is largely underutilized and (2) drug discovery in this area requires novel approaches. Below, we will attempt to summarize several factors determining the relative scarcity of pGPCR-targeted drugs.

While the structural differences between peptide ligands for pGPCRs and endogenous small molecules modulating the function of non-pGPCRs (such as biogenic amines, nucleotides and ions) may be apparent, the structural distinctions between orally active peptidomimetics acting on pGPCRs and synthetic non-pGPCR ligands are often only faintly expressed. We decided to elucidate possible differences between the synthetic ligands to pGPCRs and non-pGPCRs expressed in terms of molecular topology, polarity and hydrogen bonding potential. The results of our studies are summarized below.

Recently, we investigated the difference between several receptor-specific groups of GPCR ligands by using Kohonen self-organizing maps (SOMs) [31]. Because of some problems in the analysis of multidimensional property spaces inherent to Kohonen SOM methodology, we performed a complementary study specifically aimed at discrimination of synthetic small-molecule ligands to pGPCRs and non-pGPCRs. In theory, the algorithm of nonlinear mapping (Sammon mapping) [32] is especially attractive for data visualization and data mining, as the resulting

mapping gives an insight into the presence and structure of clusters in the data, and each projection point corresponds to a specific data entry. The general idea was to create a two-dimensional image of a multidimensional original property space. The distance between two points on the map directly reflecting the similarity of the compounds [33–35]. As input variables, we used electrotopological state (E-state) indices [36], which encode information about both the topological environment of an atom, and the electronic interactions resulting from all other atoms in the molecule. Unlike many other types of molecular descriptors, E-state indices are easily and unambiguously calculated, and, at the same time, they encode some essential molecular features characterizing the topology, polarity and hydrogen bonding capabilities of a molecule.

We collected a 593-compound database of known GPCR modulators including both drugs and compounds that had entered preclinical and clinical trials (Prous Ensemble database). Two non-overlapping data sets were included in the database: the first data set consisted of 186 pGPCR ligands, and the second included 407 agents active against non-peptidergic GPCRs. The database was filtered based on molecular weight (not more than 600) to ensure its small-molecule status. A total of 24 E-state indices were calculated for each molecule.

Figure 8.3 shows the distributions of the compound categories studied on the Sammon map. Small-molecule pGPCR ligands are shown as black circles, and non-pGPCR active drugs are indicated by white circles. There are clear differences in their location. The structures shown and the locations of three launched pGPCR (on the left) and three non-pGPCR (on the right) active drugs provide some visual clues for their discrimination. The differences are particularly evident for olmesartan medoxomil and desloratadine which occupy diametrally opposed positions on the map. In our assessment, pGPCR ligands are topologically more complex and have an increased number of polar functional groups.

This method provides a reasonable basis for the assessment of the pGPCR activity potential. For example, it can be used as a computationally inexpensive virtual filtering procedure in the design of pGPCR-targeted libraries. The exercise described also illustrates well the increased complexity of synthetic pGPCR ligands, expressed in terms of atomic electrotopological state, as compared to non-pGPCR ligands.

We believe that difficulties in elucidation of a specific interaction pattern between a pGPCR ligand and a respective receptor arise for several reasons. First, there is great diversity of endogenous ligands capable of activating various pGPCRs. The 20 amino acids used to build proteins and peptides derived from genes provide enormous chemical diversity, as well as H-bond-accepting and donating properties, hydrophobicity, conformational space, etc. There are astronomical numbers of possible structures for even a small protein: thus, $\sim 10^{13}$ structures are theoretically possible for decapeptide, and posttranslational modifications introduce additional complexity. Furthermore, a single peptide or protein can assume different conformations (ϕ/ψ space) and many topographies (χ space). There is a high dynamic diversity of peptides and proteins that can readily change conformations and topographies depending on temperature, ionic strength and
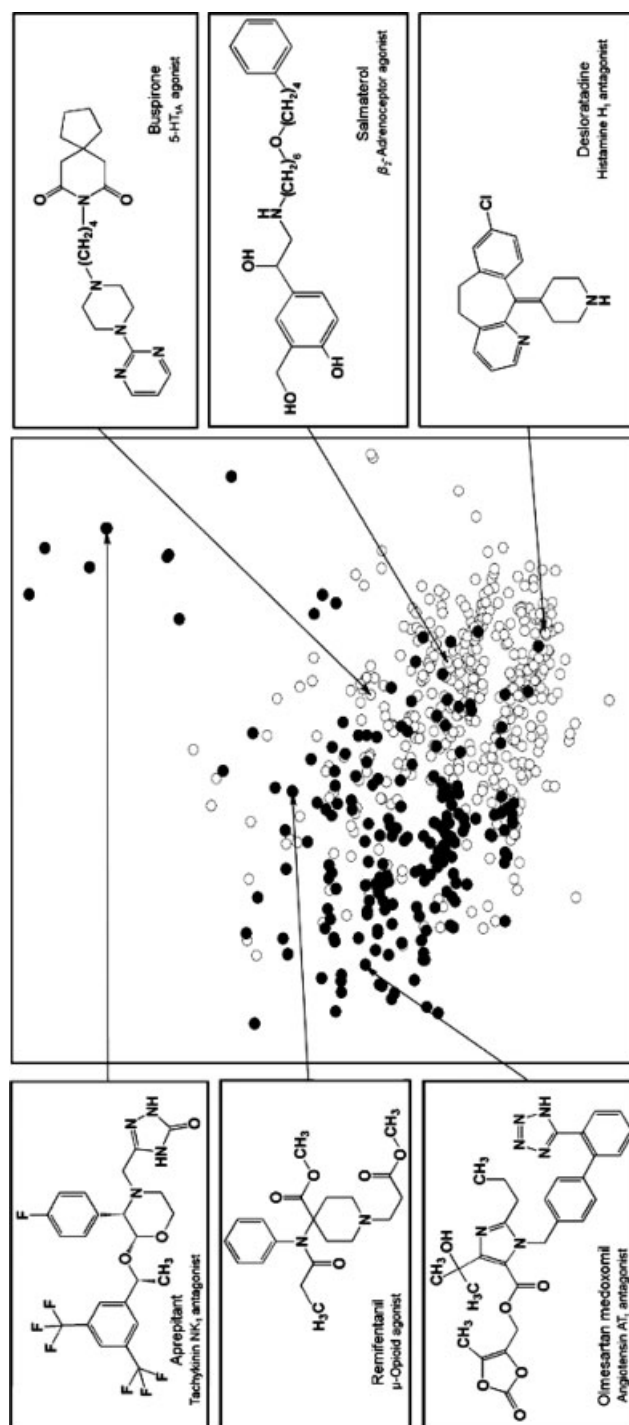
Fig. 8.3 Nonlinear map illustrating the differences between small-molecule synthetic ligands to peptidergic (black circles) and non-peptidergic (white circles) GPCRs expressed in terms of atomic electrotopological state.

microenvironmental effects. To further complicate matters, there is the aforementioned evidence for the possible diversity of peptides available to activate a particular pGPCR, the induced fit upon ligand binding and cooperative effects.

The problems mentioned suggest that there are many relevant peptide ligand–receptor interaction patterns, which make it problematic to achieve success when using a single approach to pGPCR library design. The observed increased complexity of pGPCR drugs as compared to non-pGPCR agents provides further evidence of the difficulties associated with pGPCR-directed drug discovery.

Despite the obstacles mentioned above, the track record of success for this target family suggests the great potential benefits associated with therapeutic intervention at these receptors. Nowadays, drug discovery directed towards pGPCR ligands has achieved success through integration of genome data analysis, virtual and high-throughput screening, combinatorial chemistry and classical medicinal chemistry. In the following sections, we will describe several successful pGPCR-directed lead discovery programs with emphasis on the underlying discovery concepts.

8.2.2
Docking and Pharmacophore-based Design

Due to the rapidly increasing availability of structures of the target proteins or homology models, which can be used as templates for virtual screening, the protein structure-based methods are becoming increasingly popular in the field of target-directed library design. Today, we are witness to a clear shift from the ligand-structure-based methods towards more and more sophisticated docking algorithms [37]. Information concerning the primary structure of GPCRs has become increasingly accessible from their sequencing, and 3D molecular models can be developed based upon a homologous X-ray structure [38, 39].

There are several studies, which demonstrated the potential of docking methods for the discovery of small molecule pGPCRs ligands. For example, Underwood and co-workers generated a 3D model of the human AT1 receptor based on bacteriorhodopsin and were able to dock the losartan-type antagonist L-158,282 (MK-966) into the biogenic amine binding site [40]. Another study has been presented which applies structural models of chemokine and biogenic amine receptors within the chemical optimization program of a CCR5 lead [41]. The initial lead compound showed high affinity towards the CCR5 receptor.

Despite the definite success in the lead discovery programs described, practical utility of the target-structure-based approach in the screening of large virtual libraries directed towards pGPCRs is still limited because of the lack of quality crystallographic data, detailed knowledge of the ligand binding mode and inherent issues concerning scoring functions. To further complicate matters, there are some specific concerns associated with pGPCRs as mentioned in the preceding sections. In most cases, the hit rate of virtual screening approaches benefits from the availability of potent, structurally diverse, and conformationally restricted receptor ligands as starting points.

The natural ligand can provide a good starting point in the lead-finding process. In this case, a valid pharmacophore model can be generated and then employed to identify lead structures with novel scaffolds. Hruby et al. designed a non-peptide ligand for the δ opioid receptor on the basis of a conformationally constrained peptide lead [42] (Fig. 8.4 a). Extensive computation studies led to the determination of the three-dimensional disposition of the key pharmacophore units. Then a variety of non-peptide scaffolds were explored on which these key pharmacophore elements were placed. As a validation of this approach, several highly potent and δ-selective non-peptide ligands were identified as exemplified by 1,4-piperazine analog **1** [43].

The researchers at Merck demonstrated another successful application of the 3D pharmacophore technique to identify novel non-peptide lead structures (Fig. 8.4 b). Optimization of somatostatin-derived peptides has resulted in the cyclic hexapeptide L-363,377 as a somatostatin agonist [44]. The side chains of the Tyr-D-Trp-Lys motif required for the biological activity were used as a probe to search Merck's compound collection with a 3D search method [45]. Biological testing of compounds selected by using the 3D pharmacophore search yielded com-
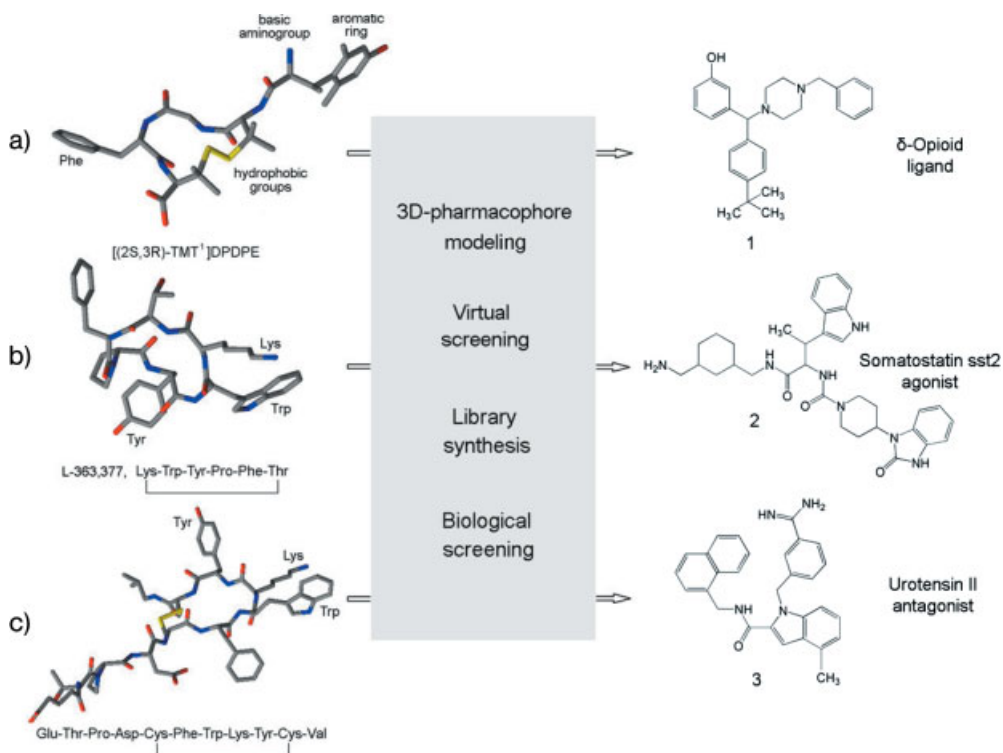
Fig. 8.4 Application of 3D pharmacophore technique for identification of non-peptide pGPCR ligands.

pounds which displayed high inhibitory activity against the human SSTR2 receptor. Combinatorial follow-up synthesis and screening led to the discovery of highly active agonist for the human SST2 receptor (structure **2**) as well as other subtype-selective somatostatin agonists.

Recently, researchers at Aventis have employed a similar virtual screening strategy to identify non-peptide antagonists for the urotensin II receptor [46] (Fig. 8.4c). The spatial arrangement of the Trp-Lys-Tyr motif identified as the important pharmacophoric pattern was deduced from the NMR spectroscopy solution structure of human urotensin II and of the disulfide bridged analog Ac-Cys-Phe-D-Trp-Lys-Tyr-Cys-NH$_2$. These results were translated into respective 3D pharmacophore models, which were used to virtually screen the Aventis compound collection. By using the pharmacophore model of the human urotensin II structure as a query, 10 of the 500 virtual hits (2%) showed biological activity with an IC$_{50}$ value of 400 nM for the best compound **3**. The virtual screening approach has identified numerous novel scaffolds with reasonable antagonist activity providing promising starting points for subsequent chemical optimization programs.

As an extension of this discovery approach, Mason and co-workers have described a four-point pharmacophore method for molecular similarity, which calculates all potential pharmacophores for a given molecule [47]. The authors were able to identify GPCR-specific pharmacophores and to apply the pharmacophore key to the design of GPCR-focused libraries.

8.2.3
Knowledge-based Data Mining Approaches

Knowledge-based data mining methods used for correlation of molecular properties with specific activities can be useful in the discovery of pGPCR-targeted leads. For example, the recursive partitioning (RP) algorithm, which uses decision trees and binary descriptors to identify specific partitions enriched with active molecules, was used for analysis of a large number of μ-opioid receptor ligands [48]. It has been shown that the optimal RP model has been able to "discover" the existence of the two main ("morphine-like" and "meperidine-like") μ ligand families, represented as the two main "active nodes" of this tree. Several groups reported application of artificial neural networks [49, 50] and Kohonen self-organizing maps (SOMs) [31, 51] for the design of GPCR-specific libraries. These methods provided a reasonable basis for the computationally inexpensive assessment of the GPCR activity profile.

One recent study demonstrates the high potential of the virtual screening strategy based on a data mining method for enhancement of productivity of pGPCR-directed lead discovery [52]. The study was focused on GPCRs that are activated by positively charged peptide (GPCR-PA$^+$) ligands. Using a special partitioning algorithm based on five calculated molecular descriptors, a region of chemical property space enriched in GPCR ligands was identified. This information was used to design and synthesize a "test" library of 2025 single, pure compounds to sample

portions of this property space associated with GPCR-PA$^+$ ligands. The library was evaluated by high-throughput screening against three different pGPCRs, rMCH, hMC4, and hGnRH, and found to be highly enriched in active ligands (4.5–61-fold) compared to a control set of 2024 randomly selected compounds.

8.2.4
Chemogenomics Approaches

The effective identification of high quality hits and leads across diverse classes of GPCR targets can be based on the systematic analysis of structural genomics data [53]. The latest human genome initiatives allow for establishing the relationships between ligands and targets, and thus, offer the potential to use the knowledge obtained in the screening experiments of one target in lead finding for another. Several approaches to explore the chemogenomics knowledge space were recently reported together with their use for the generation of target-directed libraries. The key element of this knowledge space is the ligand–target matrix, which represents a comprehensive data source suitable for effective data mining. The collection of properly annotated ligand–target databases can help to elucidate the mechanism and evaluate potential target specificity of small molecule ligands [54].

Researchers at Novartis described an annotation scheme for GPCR ligands for *in silico* screening and combinatorial design of targeted libraries [55, 56]. Retrospective *in silico* screening experiments have shown that such reference sets can be useful for the identification of ligands binding to receptors closely related to the reference system. Such a systematic exploration of the ligand–target matrix for selected target families appears to be a promising method of accelerating pGPCR-directed drug discovery. A prominent chemogenomic approach to the design of GPCR-targeted libraries has been developed by scientists at BioFocus [30]. Thematic Analysis defines a common consensus binding site for all GPCRs in the upper half of the transmembrane region. Commonly occurring small sets of amino acids (themes) are identified from primary sequence overlays and associated with ligand fragments (motifs) using SAR information. Multiple themes have been identified across the GPCR family and these have been associated with motifs to create a design tool for combinatorial libraries and lead optimization.

8.2.5
Incorporation of Specific Biomolecular Recognition Motifs

Identification of structural patterns in biomolecular recognition and their incorporation into the structures of small molecules is the key to the design of novel drugs. However, our ability to design the desired molecular structures and plan the reaction pathways currently lacks guidance from our knowledge of small molecule-binding sites on biological macromolecules. This is particularly true for peptidergic GPCRs. In this situation, the knowledge of some common structural elements of biomolecular recognition can be used to constrain the structures of synthetic compounds to those optimally fitted for binding.

The following sections present several useful strategies which were applied in the design of focused chemical libraries targeting pGPCRs. The strategies are based on information regarding specific ligand-receptor recognition motifs such as privileged structures, β-turns and other elements of peptide mimicry, which can be used for the rational design of a library.

8.2.5.1 Privileged Structures

The mining of large databases with experimental activity data led to identification of specific chemical fragments that appear over-represented among high affinity ligands to multiple, unrelated classes of protein receptors [57]. Such fragments are commonly known as "privileged structures" and can be an essential tool in targeted GPCR library design [58–61].

Privileged substructures can be found in ligands over a broad set of GPCRs. It was demonstrated that the privileged structures recognize a conserved binding pocket in a subset of GCPRs and that this "common" binding domain is complementary to the privileged structure [62]. In order to obtain evidence for the existence of privileged substructures over a wide number of pGPCR families, we performed analysis of the frequency of occurrence of several typical GPCR privileged structures taken from literature [60–63] in two large databases of pGPCR and non-pGPCR agents (5231 and 7811 compounds, respectively). The structures were obtained from the Prous Ensemble database. The results of our calculations are shown in Table 8.2. To quantitatively assess an enrichment of a particular pGPCR family with a specific fragment, we used the enrichment factor (F_e) calculated as the percentage of this fragment in a particular pGPCR set relative to its percentage in the entire non-pGPCR data set. The results obtained demonstrate the presence of certain receptor-specific preferences for particular substructures in the databases studied. In several cases, enrichment factors of 5–10 are observed. Extreme F_e values obtained for benzazepinone/cholecystokinin and biphenyl/angiotensin pairs (12.14 and 18.19, respectively) reflect the existence of "superprivileged" binding patterns for these fragments and the corresponding receptors. The third column of Table 8.2 shows a general measure of the irregularity of distribution of these fragments between the entire pGPCR and non-pGPCR databases.

This information can be useful for the selection of building blocks during the design of pGPCR-targeted libraries.

Examples of ligands that share the same privileged motif include the neurokinin NK2/NK3 antagonist **4** [64], the growth hormone secretagogue **5** [65], and the melanocortin subtype-4 receptor agonist **6** [66] as shown in Fig. 8.5. Despite the fact that compounds **4–6** all possess the same privileged structure, their receptor selectivity differs, presumably as a result of the influence of the capped amino acid and dipeptide segments. The spiro-piperidine-indane core of these compounds is also found in oxytocin, somatostatin, tachykinin, melanocortin, and anaphylatoxin chemotactic receptor ligands, and represents a typical privileged structure.

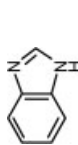**Table 8.2** pGPCR versus non-pGPCR receptor-specific preferences for particular substructures.[a]
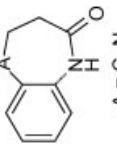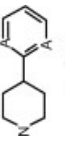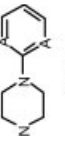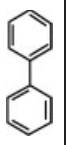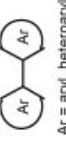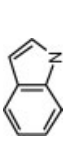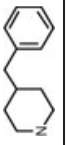
No.	Fragment	Entire pGPCR data set (5231 compounds) $\%/F_e$	Angiotensin (1555 compounds) $\%/F_e$	Opioid (513 compounds) $\%/F_e$	Endothelin (731 compounds) $\%/F_e$	Tachykinin (1044 compounds) $\%/F_e$	Neuropeptide Y (391 compounds) $\%/F_e$	Cholecystokinin (266 compounds) $\%/F_e$	Chemokine (318 compounds) $\%/F_e$
1	A = C, N	3.46/ 0.83	0.13/ 0.03	–	2.74/ 0.66	10.77/ 0.18	2.56/ 0.61	50.75/ 12.14	1.26/ 0.3
2		7.88/ 3.02	0.96/ 0.37	17.54/ 6.72	8.84/ 3.39	16.66/ 6.39	13.04/ 5.00	0.38/ 0.14	5.66/ 2.17
3		4.38/ 1.98	6.94/ 3.14	3.51/ 1.75	0.82/ 0.37	6.13/ 2.77	4.60/ 2.08	1.50/ 0.68	3.46/ 1.57
4		9.21/ 2.3	21.22/ 5.30	1.75/ 0.44	7.78/ 1.95	1.34/ 0.34	12.28/ 3.07	0.75/ 0.19	7.55/ 1.89
5	A = C, N, O	2.35/ 1.22	0.13/ 0.067	3.31/ 1.73	0.41/ 0.21	5.84/ 3.04	4.86/ 2.53	5.26/ 2.74	2.20/ 1.46
6	A = C, N, O	0.60/ 4.29	–	0.19/ 1.40	0.14/ 0.98	1.34/ 9.58	1.53/ 10.96	0.38/ 2.69	2.52/ 17.97

Table 8.2 (continued)

No.	Fragment	Entire pGPCR data set (5231 compounds) %/F_e	Angiotensin (1555 compounds) %/F_e	Opioid (513 compounds) %/F_e	Endothelin (731 compounds) %/F_e	Tachykinin (1044 compounds) %/F_e	Neuropeptide Y (391 compounds) %/F_e	Cholecystokinin (266 compounds) %/F_e	Chemokine (318 compounds) %/F_e
7	A = C, N	4.82/ 0.85	0.064/ 0.01	12.28/ **2.17**	–	11.4/ **2.01**	10.23/**1.81**	–	9.75/ **1.72**
8		0.78/ 0.79	–	2.73/ **2.76**	1.78/ **1.8**	0.38/ 0.38	2.3/ **2.32**	–	–
9	A = C, N	1.49/ 0.09	–	0.58/ 0.03	–	4.21/ 0.24	5.88/ 0.34	1.13/ 0.07	–
10		21.83/ **6.31**	65.47/ **18.19**	0.78/ 0.23	9.3/ **2.69**	1.92/ 0.55	2.56/ 0.74	–	4.09/ 1.18
11	Ar—Ar Ar = aryl, heteroaryl	33.66/ **2.19**	80.13/ **5.21**	4.29/ 0.28	26.13/ **1.7**	11.78/ 0.77	30.95/ **2.01**	6.39/ 0.42	10.69/ 0.7
12		0.99/ 0.07	3.09/ 0.21	1.36/ 0.09	14.09/ 0.97	19.83/ 1.36	6.65/ 0.46	17.29/ 1.19	8.81/ 0.6
13		1.01/ 0.78	–	–	–	1.34/ 1.04	0.26/ 0.2	0.38/ 0.29	11.32/ **8.78**

a) The first value in each cell shows the percentage of compounds in the particular pGPCR-specific group of ligands, which contain this substructure; the second value is the enrichment factor (F_e) for the same substructure. Empty cells indicate the absence or statistically insignificant data. F_e values higher than 1.5 are shown in boldface.
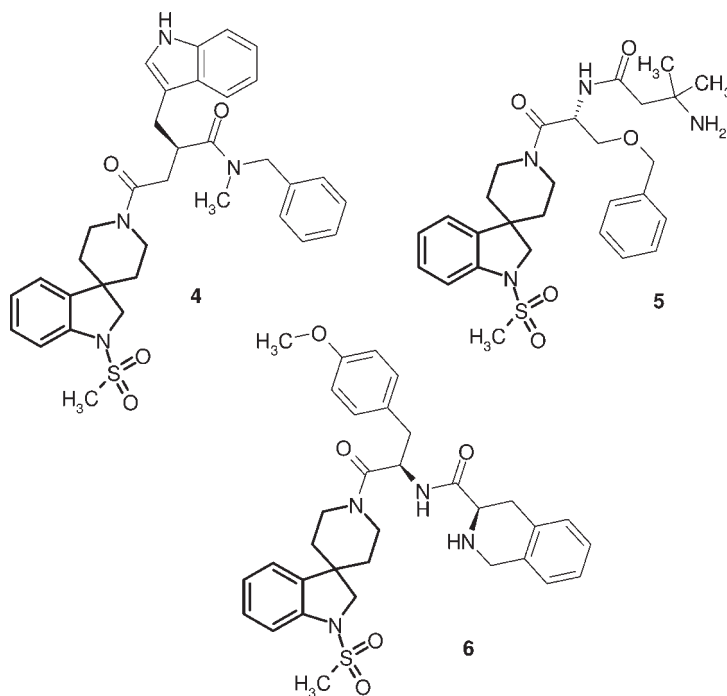
Fig. 8.5 Antagonists of three different pGPCRs containing the spiro-piperidine-indane privileged fragment (highlighted in bold).

The incorporation of privileged fragments as core or terminal building blocks has been used as a useful strategy for the generation of pGPCR-targeted libraries [31, 57–59]. One recent report described the evaluation of a wide range of structurally diverse privileged structures linked to the key dipeptide D-Trp-Lys-OMe [63]. This dipeptide is a low molecular weight SSTR2 agonist identified as having good overlap with the calculated β-turn about the Tyr-Trp-Lys-Thr sequence found in the cyclic hexapeptide SSTR agonist [45]. Privileged units such as biphenyls, 1,1-diphenylmethyls, diphenylethers, arylindoles, benzimidazolones, benzimidazoles, tetrahydroisoquinolines, etc., were incorporated as part of the amine and carboxylic acid building blocks. Despite the fact that the incorporated subunits varied substantially in size, shape, flexibility and polarity, of approximately 250 compounds evaluated, as many as 50 resulted in highly potent analogs in the murine SSTR2 binding assay, falling into 10 different types of privileged structure. It is significant that privileged structures themselves were required for potent binding, since analogous peptide and capped peptide analogs failed to produce potent ligands.

The design based on privileged structures seems to be a very useful approach for rapid identification of ligands to putative targets [67]. This design is inherently consistent with the homology-based similarity principle, which constitutes an essential part of the chemogenomic strategies. As was mentioned above, using a ser-

ies of docking experiments for a set of class A GPCRs, a good correlation has been found between conservation patterns of residues in the ligand binding pocket and the privileged structure fragments in class A GPCR ligands [62]. As a result, target-directed ligand libraries can be effectively designed without any foreknowledge of the structure of the endogenous ligand, which in turn means that even orphan receptors can be addressed as potential drug targets.

An inherent concern associated with this approach is the restricted availability of privileged substructures for known target families. The resulting issues concerning intellectual property and selectivity of action clearly limit the scope of the approaches based on privileged structures. This problem can be addressed by using special methods of rational transformation of active scaffolds aimed at maximal exploration of the chemical space surrounding the initial hits and generation of analogs with enhanced IP position [68]. The goal of this procedure is to select structures with similar steric and electrostatic parameters but different chemistry. A bioisosteric approach, which is one of the key concepts in drug design, represents a useful lead optimization strategy [69].

8.2.5.2 Mimetics of the Peptide Secondary Structure Elements

Secondary structure elements in proteins play a key role in molecular recognition events in biological systems through their characteristic three-dimensional presentation of functional groups on their surfaces. Many peptide ligand–receptor interactions are initiated or mediated by a key local secondary structural element in the protein [70]. Therefore, small molecules bearing a similar local structural feature can effectively mimic the ligand binding function of a peptide. In particular, mimicking the β- and γ-turns of natural pGPCR ligands appeared to be a very useful approach in the design of effective small-molecule peptidomimetics.

The concept of β-turn mimicry is that the vectors of the functional groups at the i + 1 and i + 2 positions and the incoming and outgoing peptide chains adopt similar positions to the corresponding vectors of an idealized β-turn (Fig. 8.6). Eguchi et al. described a β-turn peptidomimetic with four sites of diversity readily accessible through solid phase synthesis from commercially available diversity components [71, 72]. Several compounds with moderate to low nanomolar binding affinity for a non-selective opioid receptor were identified and the most potent (structure **7**) was shown to be a selective, full agonist at the μ-receptor (IC_{50} = 9 nM). Interestingly, the linear peptide corresponding to structure **7** did not demonstrate binding to the same non-selective receptor.

The synthesized bicyclic template of compound 7 was shown by X-ray crystal structural analysis and solution phase 2D-NMR spectroscopy to mimic a type I β-turn conformation [71]. This highly constrained 6,6-bicyclic system incorporating functionality at the i to i + 3 positions afforded an opportunity to probe the biologically active conformation of peptides that potentially adopt a reverse turn conformation. This scaffold was also applied to the preparation of mimetics of Leu-enkephalin to identify potent and selective ligand(s) for opioid receptors (μ, δ and κ) and to elucidate their structure–activity relationship (SAR) [71].
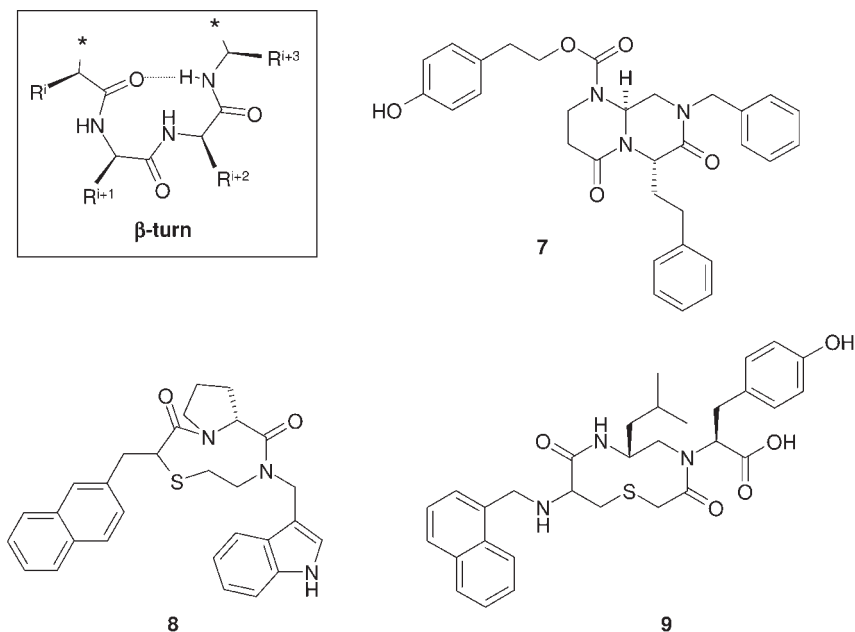
Fig. 8.6 The structure of the β-turn and its mimetics.

One of the first reported non-peptide ligands for the melanocortin receptors was based upon the β-turn [73]. The β-turn mimetics were based on the Phe-Arg-Trp peptidyl side chains that represent the putative pharmacophore of melanocortin peptides. A library of 951 β-turn compounds was screened at the MC1R for agonist activity and several compounds resulted in receptor stimulation above basal level (exemplified by structure **8**, Fig. 8.6). In a follow-up study, a novel cyclic thioether scaffold was used to mimic the β-turn [74]. A series of 19 compounds were screened for agonist activity at the mouse melanocortin receptors. Several compounds were identified with agonist activity. An example of one non-peptide melanocortin agonist **9** from this series based on the β-turn, is shown in Fig. 8.6. These studies provided some of the first non-peptide ligands for the melanocortin receptors and provided experimental evidence to support a bioactive conformation consisting of a β-turn.

Structure/conformation/activity studies of somatostatin, a 14-amino acid peptide involved in the inhibition of release of insulin, glucagon and growth hormone, had suggested that a β-turn composed of Phe^7-Trp^8-Lys^9-Thr^{10} is important for its biological activity. Based on these findings, several highly potent cyclic hexapeptide mimetics of somatostatin were designed and synthesized [75].

In a γ-turn, peptide chain reversal occurs over three residues and a hydrogen bond is formed between residues i and i + 2 so that a pseudo-seven-membered ring is formed. γ-Turns are less frequent than β-turns in biologically active peptides, but structural studies have revealed that vasopressin [76], angiotensin II [77, 78], and bradykinin [79] may adopt γ-turn conformations.
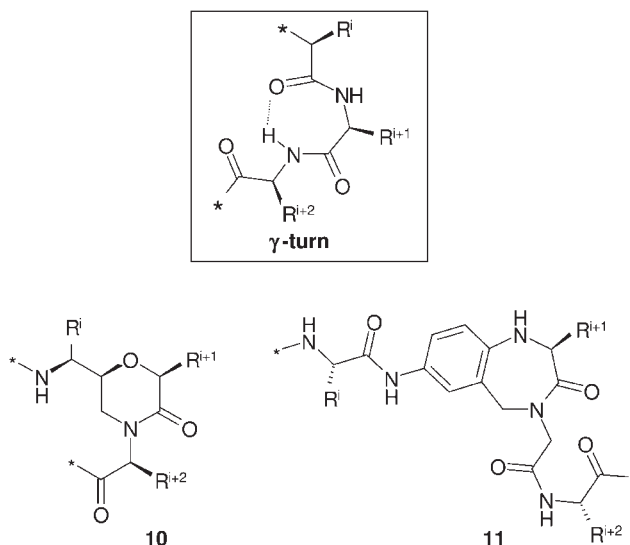
Fig. 8.7 The structure of the γ-turn and its mimetics.

Several molecular scaffolds have been reported which mimic the γ-turn (Fig. 8.7). Thus, 2,4,6-trisubstituted morpholin-3-one structure **10** successfully mimicked the γ-turns of natural ligands vasopressin [80] and oxytocin [81]. In this six-membered ring, the amide bond between residue i and i + 1 of the γ-turn has been replaced by a methylene ether isostere. In addition, a methylene bridge ensures the close spatial location of residues i and i + 2. *Ab initio* calculations revealed that this covalent linkage effectively restricts the torsional angles of the i + 1 residue to values close to those of a γ-turn [80, 82]. A selective AT2 receptor ligand with a benzodiazepine-based γ-turn-like mimetic **11** replacing the amino acid residues 4–5 of angiotensin II, was synthesized [83]. Despite these successful examples, the potential of γ-turn mimetics in the design of small-molecule therapeutics still remains to be demonstrated.

8.3
Synthesis of pGPCR-focused Libraries: Example of a Practical Methodology

In the last part of our chapter, we present one example of useful practical methodology for the design of pGPCR-targeted chemical libraries. The major difficulty that limits the effective use of small molecules in contemporary lead discovery programs is the lack of routine access to structurally complex and diverse small molecules that can be used to modulate biological systems [84]. Another, more subtle problem is the extensively advocated bias to drug-like chemistry [85]. However, the actual drug-likeness has more to do with delivery and pharmacokinetic parameters than with affinity and selectivity for a protein target. In our opinion,

identification of patterns in ligand–target recognition constitutes another reasonable basis for rationally creating small molecules that mimic or interfere with peptide ligand–receptor interactions. The basic idea of our approach is the synthesis of small molecules capable of modulation of pGPCRs, without regard for any particular receptor. To achieve this goal, an advanced parallel solution-phase synthetic strategy has been developed as an effective approach to novel diverse peptidomimetics, which contains special biomolecular recognition motifs.

This method is based on the modified multicomponent reactions compatible with high-throughput combinatorial synthesis. Multicomponent reactions can be carried out very efficiently in solution and are suitable for the synthesis of libraries of diverse small molecules. A classical example is the four-component Ugi reaction between aldehyde, amine, isonitrile and carboxylic acid, which has emerged as a powerful tool for rapid identification and optimization of lead compounds in drug discovery [86]. One important modification of this reaction that can lead to rare heterocyclic structures is the use of bifunctional reagents as starting reactants. In particular, a new Ugi-type reaction has been developed, which represents an interaction between azaheterocyclic reagents containing carboxylate and aldehyde (keto) functions in the same molecule and isonitriles and primary amines [87].

As an illustration of this methodology, the synthetic approach to various annelated 5-carbamoyl-pyrazin-3-ones starting from the bifunctional reagents containing 2-oxoethylamino-acetic acid fragments is outlined in Scheme 8.1. When compounds **13** and **16** were treated with isonitrile and primary amine in methanol, conversion into the corresponding pyrazinone libraries **14** and **17** proceeded smoothly and without any indication of major side reactions. The key bifunctional reagents **13** and **16** used in these reactions were prepared from the corresponding azaheterocyclic acids **12** or aldehydes (ketones) **15** upon their alkylation with α-brominated ketones or acids, respectively.

The scope of the synthetic approach that has been developed can be essentially extended by the synthesis of novel annelated seven-membered heterocycles
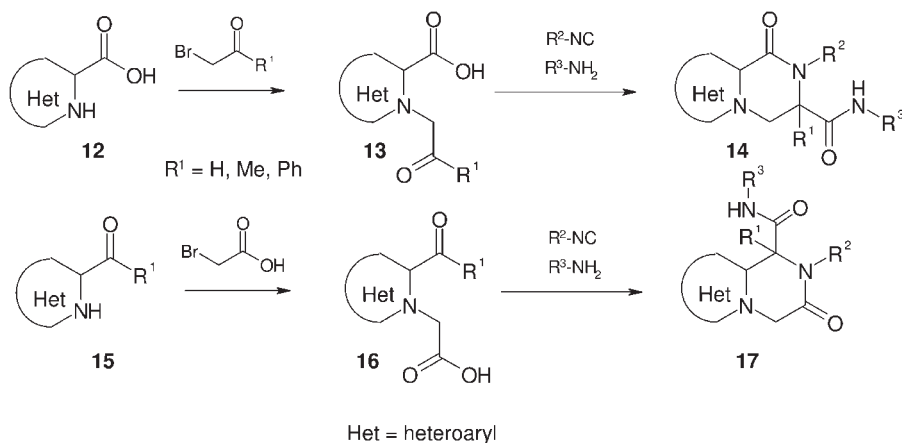
Scheme 8.1
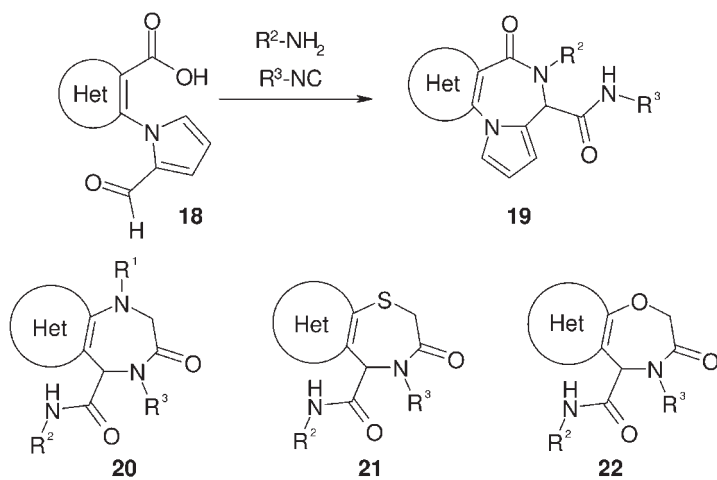
Scheme 8.2

(Scheme 8.2). For example, a series of novel pyrrolodiazepines 19 were effectively prepared when using aldehyde acids of general structure 18 as starting bifunctional reactants. A wide number of novel fused seven-membered azaheterocycles outlined in Scheme 8.2 (structures 20–22) can be prepared in an analogous manner.

The synthesized molecules illustrated in Fig. 8.8 contain a number of privileged motifs, e.g. diazepinone (structures a and b), indole, biaryl, etc. The piperazinone and azepinone scaffolds constitute examples of β- and γ-turn mimicks (c, d, f–h) with an inherent potential for combinatorial exploration of functional diversity [88]. In a more general sense, compounds synthesized by this method constitute examples of conformationally rigid peptidomimetic cyclic molecules. Thus, the piperazinone and azepinone rings can be viewed as a means to constrain the torsion angle of an amino acid's backbone bonds in a dipeptide mimic. The synthesized compounds contain certain characteristic conformationally-fixed amino acid fragments, for example: conformationally-constrained dipeptide moieties (d–g), phenylalanine-like (h) and phenylglycine-like (i) fragments.

The synthetic method is noteworthy for its ability to generate complex, diverse and little explored structures from simple building blocks. The synthesized libraries represent compact focused collections of small molecules containing special biomolecular recognition elements typical of pGPCR-targeted peptidomimetics. Finding and analyzing primary actives can eventually lead to the identification of particular protein targets capable of being modulated by these small molecules. Biological evaluation of the heterocycle-fused pyrazinones and diazepinones is currently in progress with respect to a number of pGPCR biotargets.
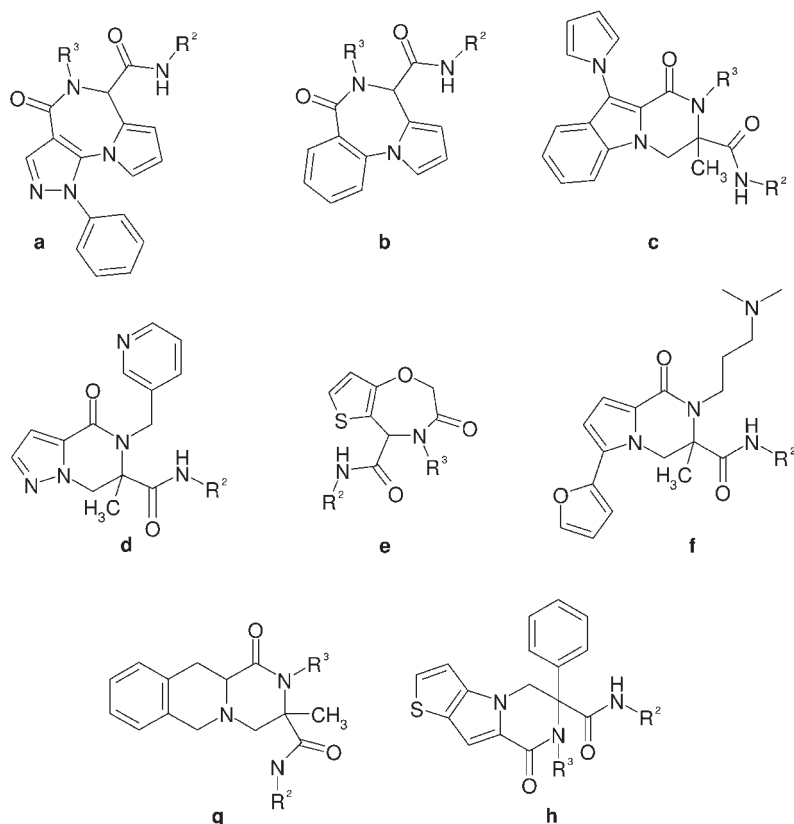
Fig. 8.8 Examples of synthesized compounds containing various biomolecular recognition motifs.

8.4
Conclusions

The study of peptidergic GPCRs and their natural and synthetic ligands is an intense area of research in drug discovery with promising clinical potential. Long-term studies in this field led to the discovery of several small-molecule agents progressed into successful marketed drugs. There has also been significant progress in the identification of orally active small-molecule nanomolar agonists and antagonists, although their clinical efficacy still remains to be demonstrated.

As outlined in this chapter, successful discovery of novel pGPCR leads relies on a combination of techniques from a wide range of disciplines, including molecular docking, pharmacophore-based design, sophisticated data mining methods, combinatorial chemistry and traditional medicinal chemistry. Improved bioactivity and selectivity are the main driving forces in the design of mimetics of natural peptide ligands. However, the non-peptidic nature of mimetics can also improve

the undesirable therapeutic characteristics of proteins or peptides, such as poor bioavailability, lack of oral activity, short duration of action, and potential antigenicity. The integration of chemogenomic knowledge-based ligand design strategies with advanced virtual screening technologies holds great promise for more efficient discovery of leads across diverse pGPCR families.

In contrast to other GPCR types, such as those that use biogenic amines or nucleosides as endogenous ligands and represent the most successful drug targets in the pharmaceutical industry [1], the market potential of pGPCRs still remains largely underutilized. However, based on the critical roles pGPCRs play in numerous aspects of human physiology, it is believed that the tremendous effort of numerous international groups of scientists working in the area will pay serious dividends for the foreseeable future.

References

1 J. Drews. Drug discovery: a historical perspective. *Science* **2000**; *287*: 1960–1964.

2 Lee, D., S. R. George, and B. F. O'Dowd. Novel G-protein-coupled receptor genes expressed in the brain: continued discovery of important therapeutic targets. *Expert Opin Ther Targets* **2002**; *6*: 185–202.

3 J. M. Baldwin. Structure and function of receptors coupled to G proteins. *Curr Opin Cell Biol* **2002**; *6*: 180–190.

4 J. Wess. G-protein-coupled receptors: molecular mechanisms involved in receptor activation and selectivity of G-protein recognition. *FASEB J* **1997**; *11*: 346–354.

5 Laburthe, M. et al. Receptors for VIP, PACAP, secretin, GRF, glucagon, GLP-1, and other members of their new family of G protein-linked receptors: structure–function relationship with special reference to the human VIP-1 receptor. *Ann NY Acad Sci* **1996**; *805*: 94–111.

6 E. Hermans. Biochemical and pharmacological control of the multiplicity of coupling at G-protein-coupled receptors. *Pharmacol Ther* **2003**; *99*: 25–44.

7 Holst, B. et al. Two active molecular phenotypes of the tachykinin NK$_1$ receptor revealed by G-protein fusions and mutagenesis. *J Biol Chem* **2001**; *276*: 19793–19799.

8 Palanche, T. et al. The neurokinin A receptor activates calcium and cAMP responses through distinct conformational states. *J Biol Chem* **2001**; *276*: 34853–34861.

9 Rashid, A. J., B. F. O'Dowd, and S. R. George. Diversity and complexity of signaling through peptidergic G protein-coupled receptors. *Endocrinology* **2004**; *145*: 2645–2652.

10 Moller, L. N. et al. Somatostatin receptors. *Biochim Biophys Acta* **2003**; *16*: 1–84.

11 T. G. Warren, and D. Shields. Expression of preprosomatostatin in heterologous cells: biosynthesis, posttranslational processing, and secretion of mature somatostatin. *Cell* **1984**; *39*: 547–555.

12 Skeggs, L. T., J. R. Kahn, and N. P. Shumway. The preparation and function of the hypertensin-converting enzyme. *J Exp Med* **1956**; *103*: 295–299.

13 Sakurai, T. et al. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **1998**; *92*: 573–585.

14 S. A. Wank. G protein-coupled receptors in gastrointestinal physiology. I. CCK receptors: an exemplary family. *Am J Physiol* **1998**; *274*: 607–613.

15 Hosoda, H. et al. Purification and characterization of rat des-Gln14-ghrelin, a second endogenous ligand for the growth hormone secretagogue receptor. *J Biol Chem* **2000**; *275*: 1995–2000.

16 Fujii, R. et al. Identification of a neuro-peptide modified with bromine as an endogenous ligand for GPR7. *J Biol Chem* **2002**; *277*: 34010–34016.

17 O'Carroll, A. M. et al. Molecular cloning and expression of a pituitary somatostatin receptor with preferential affinity for somatostatin-28. *Mol Pharmacol* **1992**; *42*: 939–946.

18 Yamada, Y. et al. Cloning and functional characterization of a family of human and mouse somatostatin receptors expressed in brain, gastrointestinal tract, and kidney. *Proc Natl Acad Sci USA* **1992**; *89*: 251–255.

19 Arai, H. et al. Cloning and expression of a cDNA encoding an endothelin receptor. *Nature* **1990**; *348*: 730–732.

20 Sakurai, T. et al. Cloning of a cDNA encoding a non-isopeptide-selective subtype of the endothelin receptor. *Nature* **1990**; *348*: 732–735.

21 T. Kenakin. Ligand-selective receptor conformations revisited: the promise and the problem. *Trends Pharmacol Sci* **2003**; *24*: 346–354.

22 Le, M. T. et al. Angiotensin IV is a potent agonist for constitutive active human AT_1 receptors. Distinct roles of the N- and C-terminal residues of angiotensin II during AT_1 receptor activation. *J Biol Chem* **2002**; *277*: 23107–23110.

23 De Lucca, G. V. et al. Discovery and structure–activity relationship of N-(ureidoalkyl)-benzyl-piperedines as potent small molecule CC chemokine receptor-3 (CCR3) antagonists. *J Med Chem* **2002**; *45*: 3794–3804.

24 Palani, A. et al. Synthesis, SAR and biological evaluation of oximino-piperidino-piperidine amides. 1. Orally bioavailable CCR5 receptor antagonists with potent anti-HIV activity. *J Med Chem* **2002**; *45*: 3143–3160.

25 DeVita, R. J. et al. A potent, nonpeptidyl 1-H-quinolone antagonist for the gonadotropin-releasing hormone receptor. *J Med Chem* **2001**; *44*: 917–922.

26 Cooper, L. C. et al. 2-Aryl indole NK_1 receptor antagonists: optimization of indole substitution. *Bioorg Med Chem Lett* **2001**; *11*: 1233–1236.

27 Kuduk, S. D. et al. 2,3-Diaminopyridine bradykinin B_1 receptor antagonists. *J Med Chem* **2004**; *47*: 6439–6442.

28 Fotsch, C. et al. Synthesis and structure–activity relationships of trisubstituted phenyl urea derivatives as neuropeptide Y5 receptor antagonists. *J Med Chem* **2001**; *44*: 2344–2356.

29 Wess, G., M. Urmann and B. Sickenberger. Medicinal chemistry: challenges and opportunities. *Angew Chem* **2001**; *113*: 3443–3453.

30 Crossley R. The design of screening libraries targeted at G-protein coupled receptors. *Curr Top Med Chem* **2004**; *4*: 581–588.

31 Savchuk, N. P., S. E. Tkachenko, and K. V. Balakin. Rational design of GPCR-specific combinatorial libraries based on the concept of privileged substructures. In *Cheminformatics in Drug Discovery*, Oprea, T. (Ed.). Wiley-VCH: Weinheim, **2004**; 287–313.

32 J. E. Sammon. A nonlinear mapping for data–structure analysis. *IEEE Trans Comput* **1969**; *C18*: 401–409.

33 P. E. Gill, and W. Murray. Methods for a class of finite-dimensional bifurcation problems. *SIAM J Numer Anal* **1978**; *15*: 977–992.

34 Domine, D., J. Devillers, and M. Chastrette. A nonlinear map of substituent constants for selecting test series and deriving structure–activity relationships. 2. Aliphatic series. *J Med Chem* **1994**; *37*: 981–987.

35 D. K. Agrafiotis. A new method for analyzing protein sequence relationships based on Sammon maps. *Prot Sci* **1997**; *6*: 287–293.

36 L. B. Kier, and L. H. Hall. *Molecular Structure Description: The Electrotopological State*, Academic Press: San Diego, **1999**; H. H. Maw, and L. H. Hall. E-state modeling of corticosteroids binding affinity validation of model for small data set. *J Chem Inf Comput Sci* **2001**; *41*: 1248–1254.

37 Van Dogen, M. et al. Structure-based screening and design in drug discovery. *Drug Discovery Today* **2002**; *7*: 471–477.

38 Palczewski, K. et al. Crystal structure of rhodopsin A: A G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

39 J. M. Baldwin. The probable arrangement of the helices in G protein-coupled receptors. *EMBO J* **1993**; *12*: 1693–1703.

40 D. J. Underwood. Structural model of antagonist and agonist binding to the angiotensin II, AT$_1$ subtype, G protein coupled receptor. *Chem Biol* **1995**; *1*: 211–221.

41 T. Klabunde, and G. Hessler. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**; *3*: 928–944.

42 Shenderovich, M. D., Liao, S., Qian, X., and Hruby, V. J. Three-dimensional model of the δ-opioid pharmacophore: comparative molecular modeling of peptide and non-peptide ligands. *Biopolymers* **2000**; *53*: 565–580.

43 Liao, S. et al. *De novo* design, synthesis, and biological activities of high-affinity and selective non-peptide agonists of the δ-opioid receptor. *J Med Chem* **1998**; *41*: 4767–4776.

44 D. F. Veber. In *Peptides, Chemistry,* and Biology. *Proceedings of the 12th American Peptide Symposium*, Smith, J. A., Rivier, J.E. (Eds). ESCOM: Leiden, the Netherlands, **1992**; 3–14.

45 Rohrer, S. P. et al. Rapid identification of subtype-selective agonists of the somatostatin receptor through combinatorial chemistry. *Science* **1998**; *282*: 737–740.

46 Flohr, S. et al. Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure–activity relationships and nuclear magnetic resonance studies on urotensin II. *J Med Chem* **2002**; *45*: 1799–1805.

47 Mason, J. S. et al. New 4-point pharmacophore method for molecular similarity and diversity applications. *J Med Chem* **1999**; *42*: 3251–3264.

48 D. Horvath. Recursive partitioning analysis of μ-opiate receptor high throughput screening results. *SAR QSAR Environ Res* **2001**; *12*: 181–212.

49 Balakin, K. V. et al. Structure-based versus property-based approaches in the design of G-protein-coupled receptor-targeted libraries. *J Chem Inf Comput Sci* **2003**; *43*: 1553–1562.

50 Manallack, D. T. et al. Selecting screening candidates for kinase and G-protein coupled receptor targets using neural networks. *J Chem Inf Comput Sci* **2002**; *42*: 1256–1262.

51 G. Schneider, and M. Nettekoven. Ligand-based combinatorial design of selective purinergic receptor (A2A) antagonists using self-organizing maps. *J Comb Chem* **2003**; *5*: 233–237.

52 Lavrador, K. et al. A screening library for peptide activated G-protein coupled receptors. 1. The test set. *J Med Chem* **2004**; *47*: 6864–6874.

53 Jacoby, E., A. Schuffenhauer, and P. Floersheim. Chemogenomics knowledge-based strategies in drug discovery. *Drug News Perspect* **2003**; *16*: 93–102.

54 Savchuk, N., K. Balakin, and S. Tkachenko. Exploring the chemogenomic knowledge space with annotated chemical libraries. *Curr Opin Chem Biol* **2004**; *8*; 412–417.

55 E. Jacoby. A novel chemogenomics knowledge-based ligand design strategy-application to G protein-coupled receptors. *Quant Struct–Act Relat* **2001**; *20*: 115–123.

56 Schuffenhauer, A. et al. An ontology for pharmaceutical ligands and its application for *in silico* screening and library design. *J Chem Inf Comput Sci* **2002**; *42*: 947–955.

57 Evans, B. E. et al. Development of potent, selective, orally effective cholecystokinin antagonists. *J Med Chem* **1988**; *31*: 2235–2246.

58 Lowrie, J. F. et al. The different strategies for designing GPCR and kinase targeted libraries. *Comb Chem High Throughput Screen* **2004**; *7*: 495–510.

59 Lewell, X. Q. et al. RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J Chem Inf Comput Sci* **1998**; *38*: 511–522.

60 T. Guo, and D. W. Hobbs. Privileged structure-based combinatorial libraries

targeting G protein-coupled receptors. *Assay Drug Dev Technol* **2003**; *1*: 579–592.

61 HORTON, D. A., G. T. BOURNE, and M. L. SMYTHE. The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem Rev* **2003**; *103*: 893–930.

62 BONDENSGAARD, K. et al. Recognition of privileged structures by g-protein coupled receptors. *J Med Chem* **2004**; *47*: 888–899.

63 PASTERNAK, A. et al. Diverse peptidyl privileged structures as potent somatostatin receptor ligands. *Lett Drug Des Discov* **2004**; *1*: 121–125.

64 QI, H. et al. L-Tryptophan urea amides as NK_1/NK_2 dual antagonists. *Bioorg Med Chem Lett* **1998**; *8*: 2259–2262.

65 PATCHETT, A. A. et al. Design and biological activities of L-163, 191 (MK-0677): a potent and orally active growth hormone secretagogue. *Proc Natl Acad Sci USA* **1995**; *92*: 7001–7005.

66 NARGUND, R. P. et al. PCT Patent Publication, WO9964002. *Chem Abstr* **2000**; *132*: 022957p.

67 G. MULLER. Medicinal chemistry of target family-directed master keys. *Drug Discovery Today* **2003**; *8*: 681–691.

68 TKACHENKO, S. E. et al. Efficient optimization strategy for marginal hits active against *abl* tyrosine kinases. *Curr Drug Discov Techn* **2004**; *1*: 201–210.

69 X. CHEN, and W. WANG. The use of bioisosteric groups in lead optimization. *Ann Rep Med Chem* **2003**; *38*: 333–346.

70 ROSE, G. D., L. M. GIERASCH, and J. A. SMITH. Turns in peptides and proteins. *Adv Protein Chem* **1985**; *37*: 1–109.

71 EGUCHI, M. et al. Solid-phase synthesis and structural analysis of bicyclic β-turn mimetics incorporating functionality at the *i* to *i* + 3 positions. *J Am Chem Soc* **1999**; *121*: 12204–12205.

72 EGUCHI, M. et al. Design, synthesis and evaluation of opioid analogues with non-peptidic β-turn scaffold: enkephalin and endomorphin mimetics. *J Med Chem* **2002**; *25*: 1395–1398.

73 HASKELL-LUEVANO, C. et al. Compounds that activate the mouse melanocortin-1 receptor identified by screening a small

molecule library based upon the beta-turn. *J Med Chem* **1999**; *42*: 4380–4387.

74 BONDEBJERG, J. et al. A solid-phase approach to mouse melanocortin receptor agonists derived from a novel thioether cyclized peptidomimetic scaffold. *J Am Chem Soc* **2002**; *124*: 11046–11055.

75 R. M. FREIDINGER. Design and synthesis of novel bioactive peptides and peptidomimetics. *J Med Chem* **2003**; *46*: 5553–5566.

76 LANGS, D. A. et al. Structure of pressinoic acid: the cyclic moiety of vasopressin. *Science* **1986**; *232*: 1240–1242.

77 SCHMIDT, B. et al. Design, synthesis, and biological activities of four angiotensin II receptor ligands with γ-turn mimetics replacing amino acid residues 3–5. *J Med Chem* **1997**; *40*: 903–919.

78 LINDMAN, S. et al. Comparison of three ā-turn mimetic scaffolds incorporated into angiotensin II. *Bioorg Med Chem* **2000**; *8*: 2375–2383.

79 SATO, M. et al. Design, synthesis and conformational analysis of γ-turn peptide mimetics of bradykinin. *Biochem Biophys Res Commun* **1992**; *187*: 999–1006.

80 BRICKMANN, K. et al. Synthesis of conformationally restricted mimetics of γ-turns and incorporation into desmopressin, an analogue of the peptide hormone vasopressin. *Chem Eur J* **1999**; *5*: 2241–2253.

81 YUAN, Z. et al. Synthesis and pharmacological evaluation of an analogue of the peptide hormone oxytocin that contains a mimetic of an inverse ā-turn. *J Med Chem* **2002**; *45*: 2512–2519.

82 BRICKMANN, K. et al. An approach to enantiomerically pure inverse γ-turn mimetics for use in solid-phase synthesis. *Tetrahedron Lett* **1997**; *38*: 3651–3654.

83 ROSENSTRÖM, U. et al. A selective AT_2 receptor ligand with a γ-turn-like mimetic replacing the amino acid residues 4–5 of angiotensin II. *J Med Chem* **2004**; *47*: 859–870.

84 S. L. SCHREIBER. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**; *287*: 1964–1969.

85 D. E. CLARK, and S. D. PICKETT. Computational methods for the prediction of "drug-likeness". *Drug Discovery Today* **2000**; *5*: 49–58.

86 DOMLING, and I. UGI. Multicomponent reactions with isocyanides. *Angew Chem* **2000**; *112*: 3300–3344.

87 ILYN, A. P. et al. New four-component Ugi-type reaction. Synthesis of heterocyclic structures containing pyrrolo[1,2-α] [1,4]diazepine fragment. *J Org Chem* **2005**; *70*: 1478–1481.

88 GOLEBIOWSKI, A. et al. Solid-supported synthesis of putative peptide β-turn mimetics via Ugi reaction for diketopiperazine formation. *J Comb Chem* **2002**; *4*: 584–590.

9
Ligand-based Rational Design: Virtual Screening

David E. Clark and Christopher Higgs

9.1
Introduction

G protein-coupled receptors (GPCRs) constitute one of the most important target classes in pharmaceutical research, both from an historic and a current point of view. Of the top 50 products marketed in 2001, greater than 30% were targeted at GPCRs [1] and it is estimated that a similar percentage of ongoing drug discovery programs in pharmaceutical companies continue to focus on this family of receptors [2]. It is thus not surprising that a wide variety of drug design strategies have been applied in the search for new GPCR ligands [2, 3].

The application of computer-aided drug design techniques in support of GPCR drug discovery projects has received a significant boost in recent years, following the publication in 2000 of the first high-resolution X-ray crystal structure of a true GPCR (bovine rhodopsin) [4]. Even prior to this, molecular models of GPCRs had been an invaluable tool in generating ideas, rationalizing experimental data and providing a structural framework upon which to assess sometimes contradictory results [5]. Indeed, a comparison of the rhodopsin structure with the results of earlier sequence analysis and molecular modeling incorporating experimental data demonstrated a high degree of success for these methods in predicting key features of GPCR structures [6]. Nonetheless, the availability of structures of rhodopsin has provided a firmer foundation for the construction of homology models of GPCRs, particularly for amine receptors [7]. This in turn has led to some of the first successes in using GPCR homology models for virtual screening (see also Chapter 12) [8–10].

In all the excitement of recent years over structure-based virtual screening [11], it is perhaps the case that ligand-based virtual screening approaches for hit-finding against GPCRs have been somewhat overshadowed. It is thus worth pointing out that, despite recent advances, structure-based virtual screening approaches against GPCRs do not constitute a panacea and thus, ligand-based approaches should not be ignored.

9.2
Why Use Ligand-based Virtual Screening?

There are several reasons why ligand-based virtual screening presents a valuable approach to hit-finding against GPCRs.

9.2.1
Speed

The construction of a high-quality homology model of a GPCR is often time-consuming (for a recent example, see [12]). Ideally, such a model will incorporate as much experimental information as possible, in particular that provided by site-directed mutagenesis experiments. The construction and validation of a homology model is likely to be an iterative process involving successive rounds of interactive graphical inspection and refinement by molecular dynamics/mechanics. The same is likely to be true of *de novo* GPCR models, e.g. those generated by PREDICT [13]. Thus, for a project seeking to find hits against a given GPCR target, there is likely to be some time lag before structure-based virtual screening can be employed. By contrast, ligand-based approaches can be applied almost immediately following the identification of a suitable set of known ligands for the target. Additionally, ligand-based searches are at least as fast as structure-based virtual screening and, in some cases (e.g. 2-D similarity searches), orders of magnitude faster.

9.2.2
Applicability

In the available X-ray structures of rhodopsin, the receptor is in complex with retinal, an inverse agonist and thus the receptor is in an inactive conformation. It is therefore accepted that the structure is best suited to modeling inverse agonist and neutral antagonist-bound forms of GPCRs [14]. Significant challenges remain if one is seeking to create a homology model of an agonist-bound GPCR for use in virtual screening, although progress is being made in this field [15]. *De novo* structure prediction is likely to be more useful in this instance [14], although this is less widely practised than homology modeling. Once again, ligand-based techniques suffer from no such problems – all that is required is one or more known agonists on which to base searches.

9.2.3
Complementarity

Even when a high-quality homology, or *de novo*, model of the target of interest is available, there is still value in carrying out ligand-based searches. Each kind of search will yield a different set of candidate structures for evaluation prior to screening [16] and so it is perhaps most helpful to view ligand-based techniques as complementary to, rather than in competition with, structure-based approaches.

9.3
Overview of Ligand-based Virtual Screening

Successful application of ligand-based virtual screening is dependent on the availability of three components: one or more starting points, i. e. ligands suitable for use as queries, one or more databases of chemical structures, and one or more search techniques to match the query ligands against the databases.

9.3.1
Starting Points

Perhaps the most significant drawback of ligand-based approaches is that, by definition, they require knowledge of one or more known, and preferably potent, ligands for the receptor in question. In many cases, this will not present a problem and structures of active compounds may be available from in-house screening or from the literature. However, there will be occasions, perhaps for a proprietary target, when no known ligands are available or the available ligands are not readily amenable to use as starting points, e. g. large peptides. In these cases, it is not possible to apply ligand-based searching and recourse must be made to structure-based virtual screening or biochemical screening.

9.3.2
Chemical Structure Databases

Virtual screening also requires structure-searchable databases of chemical structures from which potential hits can be identified. These databases may be *actual*, i. e. comprise compounds for which physical samples exist, or *virtual*, i. e. consist of compounds not currently in existence, but which should, ideally, be amenable to rapid synthesis. In practice, this often equates to *virtual libraries* of compounds that can be readily realized by parallel or combinatorial synthesis [17].

For many companies, the foremost actual collection will be the company's corporate repository, which will often have been built up over several decades and number in the hundreds of thousands. Such collections have the advantage of containing a large percentage of proprietary molecules, but can be disadvantaged by problems arising from compound sample deterioration over time and from biases towards historic therapeutic programs or compound classes. An additional, or alternative, source of compounds is the growing number of vendors selling samples for screening. For the purposes of virtual screening, it is relatively straightforward to compile a database of over 2 million drug-like structures from the electronic catalogs supplied by these vendors [18]. The value of such databases is enhanced if they are updated regularly so that the likelihood of samples of interest being available is maximized.

While most databases will begin their lives as collections of 2-D chemical structures (often in MDL's SD file format), it is almost inevitable that a 3-D database will be required for virtual screening. The move into 3-D space allows for a greater level

of abstraction in the search and enhances the likelihood of successfully scaffold-hopping from one structural class to another. Over the last decade, the process of 2-D to 3-D structure conversion has become routine and rapid [19, 20], thanks to widely used programs such as Concord [21] and CORINA [22], which are capable of converting almost 100 % of small organic chemical structures from 2-D to 3-D.

In preparing a high-quality database for virtual screening, a number of issues need to be considered [23]. For instance, are all compounds represented in a consistent valence-bond representation? A classic example where more than one valence-bond representation is frequently used is the nitro group, which is often drawn in two ways (Fig. 9.1). Similar consideration needs to be given to the representation of groups that can exist in tautomeric forms, e.g. the hydroxypyridine and pyridone tautomers shown in Fig. 9.1. The likely protonation state of ionizable groups should also be considered. This is relatively simple for strong acids and bases, but less so for their weaker counterparts. Finally, when building a 3-D database, a decision needs to be made concerning how to handle unspecified chiral centers. It is possible to instruct some 3-D model builders to enumerate all possibilities for such centers, but this may become an unrealistic option for compounds with more than a few unspecified stereocenters.
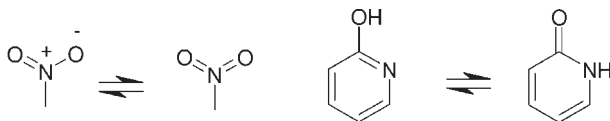
Fig. 9.1 Two valence-bond representations of the nitro group (left) and hydroxypyridine-pyridone tautomerization.

9.3.3
Database Search Techniques

Given one or more query compounds and a database of real or virtual compounds containing 2-D and 3-D chemical structures, the final part of the virtual screening jigsaw is the search technique for finding structures in the database that resemble the queries in some sense and thus, may also possess the desired biological activity. There are several possible types of search, and all may have utility in the search for novel actives [16]. In what follows, a brief introduction to the main types of database search is given. More detailed descriptions are given in the references cited in each section. Some commercially available systems for database searching are listed in Table 9.1.

9.3.3.1 **2-D Substructure Searching**
Perhaps the most straightforward type of search, both conceptually and practically, is 2-D substructure searching. This type of search operates on 2-D chemical structures and uses graph theoretical algorithms to identify database compounds that

Table 9.1 Some commercially available systems for database searching.

	2-D Substructure	2-D Similarity	3-D Substructure	3-D Similarity	Pharmaco-phore
Unity [24]	×	×	×		×
ISIS [25]	×	×	×		
Catalyst [26]			×	× (shape)	×
FlexS [27]				×	
ROCS [28]				× (shape)	
FeatureTrees [29]		×			
MOE [30]					×
Daylight [31]	×	×			

contain a query substructure [32]. These searches are fast and unambiguous – a database compound either matches the query, or it does not. 2-D substructure searches are useful for identifying other compounds in a database that share a substructure with a known active, particularly if it is thought that substructure might be (partially) responsible for the observed activity. However, some SAR data are necessary to make this kind of judgement and it may well be in the early stages of a project that only a single active in a given structural class is known. In this situation, 2-D substructure searching is unlikely to be very helpful.

9.3.3.2 2-D Similarity Searching

Substructure searches are Boolean in nature, i. e. a compound either contains the query structure (and is returned as a hit), or it does not. By contrast, similarity searches are essentially ranking procedures, ordering a database by some measure of similarity to the query and allowing the user then to browse through the highest ranking compounds [33]. As the name suggests, a 2-D similarity search seeks to identify compounds in a database that are similar to the query in terms of 2-D chemical structure. Often, these searches employ a bitstring representation of chemical structure, in which a 1 or a 0 at a given position indicates the presence or absence of a particular functional group or molecular descriptor. By comparing the bitstrings of two structures using an appropriate similarity measure, a similarity value can be computed. Typically, such values range from 0 (no similarity) to 1 (identical).

By their nature, 2-D similarity searches tend to rank compounds highly that are, in medicinal chemistry terms, probably quite close analogs of the query. In this respect, this search type is perhaps not very likely to yield novel scaffolds, although there is always a chance of locating a compound that finds a hole in a competitor's patent. The searches are useful, however, for finding analogs that can then be tested to help establish SAR.

9.3.3.3 **3-D Substructure Searching**

3-D substructure searching extends a 2-D substructure search into the third dimension, considering not only the presence or absence of one of more query substructures in a database compound, but also their relative disposition in 3-D space. Thus, in 3-D substructure searching, in order for a database compound to be considered as a hit, it must not only possess the appropriate substructures, but the distances between them must also be able to lie within specified distance ranges [34, 35]. Historically, 3-D databases stored only one conformation for each compound, but modern 3-D substructure searching systems are able to maximize the number of compounds that can be retrieved as hits by either carrying out a rapid torsional exploration of each compound that possesses the appropriate substructures to see if it can match the query distance ranges, or searching multiple stored conformations for each compound.

As might be imagined, these searches are more computationally demanding than those in 2-D, particularly if a large number of compounds need to undergo the relatively time-consuming conformational exploration step. The benefit, however, is that hits are more likely to contain scaffolds that are different from those of the query compound giving rise to potentially novel lead series.

9.3.3.4 **3-D Similarity Searching**

Many methods have been investigated for similarity searching in databases of 3-D chemical structures, some considering just the structure per se [36], others considering similarity in terms of the molecular fields around a compound [37], the pharmacophores it can express [38], or molecular shape [39]. Like 3-D substructure searching, this type of approach has great potential for the discovery of novel scaffolds, although the generation of the 3-D descriptors and/or the searches themselves can be time consuming.

9.3.3.5 **Pharmacophore Searching**

Pharmacophore searches are, essentially, a generalization of 3-D substructure searches in which the features in the query are generalized feature types (e. g. positive nitrogen, hydrophobe, hydrogen-bond acceptor) rather than specific substructures. Again, such searches can be time consuming, but once more the benefit is that hits are more likely to contain scaffolds that are different from those of the query compound giving rise to potentially novel lead series [40].

9.4
Successful Applications of Ligand-based Virtual Screening for GPCRs

Several successful applications of ligand-based virtual screening against GPCR targets have been reported in recent years. A selection of these is summarized in the following sections.

9.4.1
Somatostatin Agonists

Somatostatin exists in two bioactive forms, a 14-amino acid cyclic peptide and a 28-amino acid cyclic peptide, and is widely expressed in the central nervous system (CNS) and a number of different tissues including the intestinal tract and pancreas. This hormone has been shown to modulate the secretion of other peptides such as glucagon, growth hormone and insulin and so is used in the treatment of giantism, acromegaly and a variety of neoplasms. Yang et al. [41, 42] have used NMR studies and a proprietary 3-D similarity search engine to identify a novel somatostatin agonist.

NMR studies of the cyclic hexapeptide L-363,377 (c[PYdWKTF]), which exhibits a K_i value of 0.49 nM against the somatostatin type 2 receptor (sst2), were performed and identified a number of proximate distances between side chains. Molecular dynamics simulations using CHARMm [43] were performed on this cyclic peptide with these distance constraints applied to generate a suitable conformation for this hexapeptide. The position of the side chains of YdWK from the modeled hexapeptide were selected as probes for Merck's proprietary 3-D similarity search engine, SQ [44], which was used to interrogate the Merck chemical collection. This approach was extremely successful: the lead compound, shown in Fig. 9.2, was identified after testing fewer than 100 compounds and exhibited a K_i value of 122 nM. Further improvements in potency were obtained by increasing the amine chain length and incorporating a methyl ester to complete the lysine structure while replacing the spiroindenylpiperidine moiety by *N*-(4-piperidinyl)-benzimidazol-2-one to give rise to a compound with a K_i value of 0.01 nM. This final structure, L-054,522, is also shown in Fig. 9.2.

The structure of L-054,522 was compared to that of the cyclic hexapeptide L-363,377. This overlay revealed that the *N*-(4-piperidinyl)benzimidazol-2-one privileged structure is close to the position of the tyrosine residue. Ariens et al. [45] had proposed that the privileged structure would take up a position in an accessory binding site based on observations from some biogenic amine antagonists.
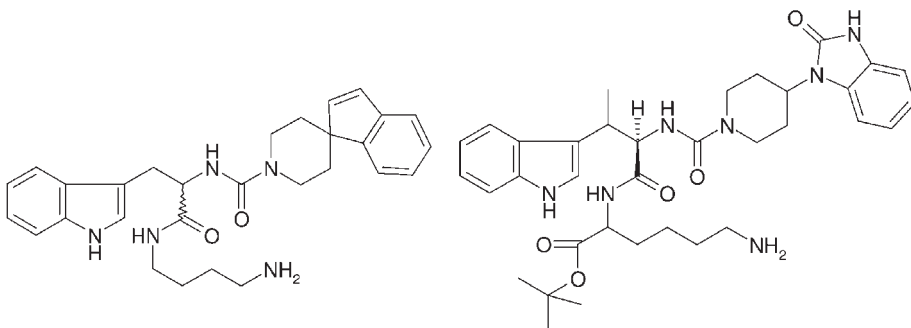
Fig. 9.2 Structure of the original hit compound identified by 3-D similarity search engine (left) and the final optimized compound, L-054,522.

However, in this case, the privileged structure of L-054,522 is located in the turn region of the peptide contradicting this hypothesis.

9.4.2
Muscarinic M₃ Receptor Antagonists

The muscarinic M_3 receptor has been shown to mediate the control of many smooth muscles, including those in the gastrointestinal tract, urinary bladder and airway. As a result, a number of muscarinic M_3 antagonists have either reached the market or are undergoing clinical trails for the treatment of diseases such as irritable bowel syndrome (Zamifenacin, Pfizer), chronic obstructive pulmonary disease (Spiriva, Boehringer Ingelheim) and overactive bladder (Darifenacin, Novartis). In the work summarized below [46], pharmacophore modeling and database searching were used to identify a number of novel lead compounds exhibiting M_3 antagonism.

Three compound series, shown in Fig. 9.3, were chosen to define the pharmacophore models. These compounds were built using Sybyl 6.2 [47] with the geometries of the fused ring system taken from X-ray crystal structures deposited in the Cambridge Structural Database [48]. A random conformational search was then performed using compound A, identifying 44 conformations that were within 5 kcal/mol of the lowest energy conformer. Compounds B and C were assumed to be rigid.
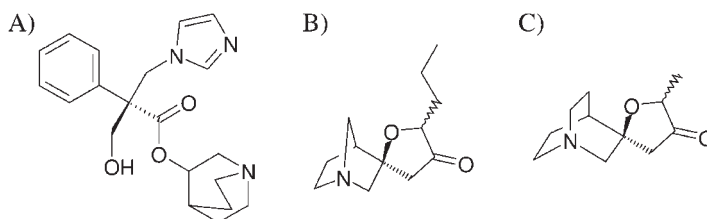
Fig. 9.3 The three-compound series used to define the pharmacophore model. (A) Lung-selective M_3 antagonist from Pfizer. (B) and (C) Compounds taken from an M_1 program displaying M_3 antagonism.

Using the standard DISCO [49] features to define the molecules (with the exception of ring centroids and the features associated with the imidazole ring in compound A as this group is believed to confer tissue specificity), a DISCO run was performed using compound B as the reference molecule. Tolerances of 0.5 Å were allowed and models were required to have between three and eight feature matches. Five models were generated, all with at least four feature matches, with two of these models showing a good overlay with the tertiary nitrogen present in the three reference molecules.

The two pharmacophore models, shown in Fig. 9.4, were used to interrogate a compound database containing compounds taken from Astra's in-house pro-
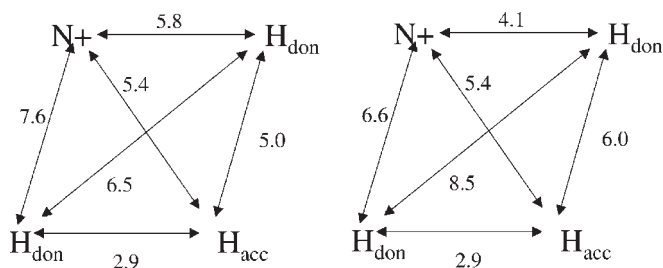
Fig. 9.4 The two pharmacophore models generated from DISCO used to interrogate the compound database. A ± 0.3 Å tolerance was applied to each distance during the search procedure.

grams and from various chemical suppliers. This approach to database construction was used in order to increase the diversity of the screening collection, because many of the compounds taken from the in-house programs were from similar chemical series. The search using the first pharmacophore model retrieved 176 compounds; that with the second pharmacophore model retrieved 173 compounds and 172 compounds were found to be common between the two sets of hits. Of the 177 compounds identified in total, 172 were screened in a tissue-based M_3 receptor assay.

The three compounds shown in Fig. 9.5 were found to possess significant activity in the M_3 receptor assay. In addition, compound D was also shown to behave as a competitive antagonist. Further assessment of these compounds revealed additional favorable properties for hit-like compounds. None of these compounds appeared to be similar to other known M_3 antagonists; physical properties such as molecular weight and logD were within acceptable ranges. Additionally, the structures appeared to be amenable to combinatorial or parallel synthesis allowing for the rapid optimization of potency and other properties.
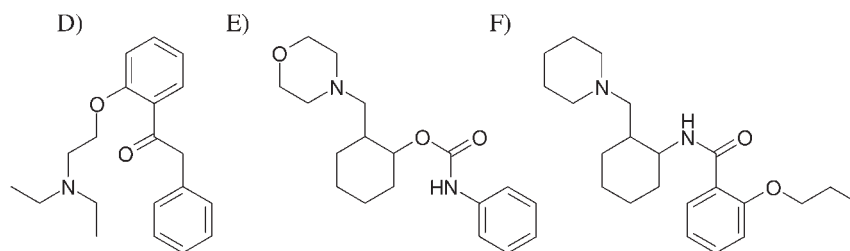
Fig. 9.5 The three compounds identified by screening using the two pharmacophore models. Compound (D) was found to be most active (pA$_2$ = 6.67 $\pm$ 0.007) while compounds (E) (pA$_2$ = 5.54 $\pm$ 0.006) and (F) (pA$_2$ = 4.83 $\pm$ 0.007) also exhibited good activity.

9.4.3
Urotensin II Antagonists

Urotensin II, an 11-amino acid cyclic peptide ETPDc[CFWKYCV], has been widely shown to act as a potent vasoconstrictor via the orphan receptor, GPR14, which is extensively expressed in cardiovascular tissues. Consequently, therapeutics targeted towards the urotensin system could be used to treat disorders including renal failure, chronic heart failure and diabetes. In fact, Actelion has Palosuran, an orally active urotensin II antagonist, in Phase IIa trials for the treatment of chronic renal failure in patients with diabetes. In the work described here, Flohr et al. [50] of Aventis performed virtual screening using a pharmacophore model, based on structure–activity relationship (SAR) data and NMR studies on urotensin II, to identify novel non-peptidic lead compounds.

The first stage of this work was to synthesize and test a number of peptides based on the urotensin II sequence in order to identify those residues responsible for activity. These experiments revealed that the minimum sequence required to retain full urotensin II activity was the cyclic hexapeptide, c[CFWKYC]. Alanine-scanning mutagenesis also confirmed that replacement of any of the exocyclic residues by alanine had no affect on potency. However, any peptides in which a residue in the cyclic hexapeptide had been replaced, with the exception of phenylalanine, exhibited an 80- to 6000-fold drop in potency and were unable to activate the receptor fully, functioning as partial agonists. Deletion of any residue from the FWKY sequence contained in the cyclic hexapeptide resulted in a considerable reduction in, or complete loss of, potency. A similar effect was observed on inverting the chirality of these residues, with the exception of tryptophan.

The SAR data revealed that the WKY sequence of urotensin II was critical for receptor activity. However, to generate a pharmacophore model, it was necessary to identify the spatial arrangement of these residues within urotensin II. NMR spectroscopic studies were performed, identifying 57 distance constraints that were used as input for a 350-ps restrained molecular dynamics simulation. The structures generated showed little deviation, with the backbone of the cyclic hexapeptide domain having an average root-mean-square (RMS) deviation of 0.45 Å. The resulting solution structure was then used as a template for a three-point pharmacophore model, incorporating the positive ionizable center from the lysine residue and two hydrophobic aromatic centers from the tryptophan and tyrosine residues. The pharmacophore query is shown in Fig. 9.6.

Conformationally restricted peptides exhibiting activity may gives clues concerning the bioactive conformation and thus the cyclic hexapeptide, Ac-c[CFWKYC]-NH$_2$, was synthesized. This exhibited sufficient activity to be used as a template for a second three-point pharmacophore model. Again, NMR spectroscopy was used, generating 32 distance constraints that were used in a 500-ps restrained molecular dynamics simulation. The structures obtained exhibited an average RMS deviation of 0.27 Å over the backbone atoms. As before, the resulting solution structure was used as a template for a three-point pharmacophore model, incorporating the positive ionizable center from the lysine residue and two
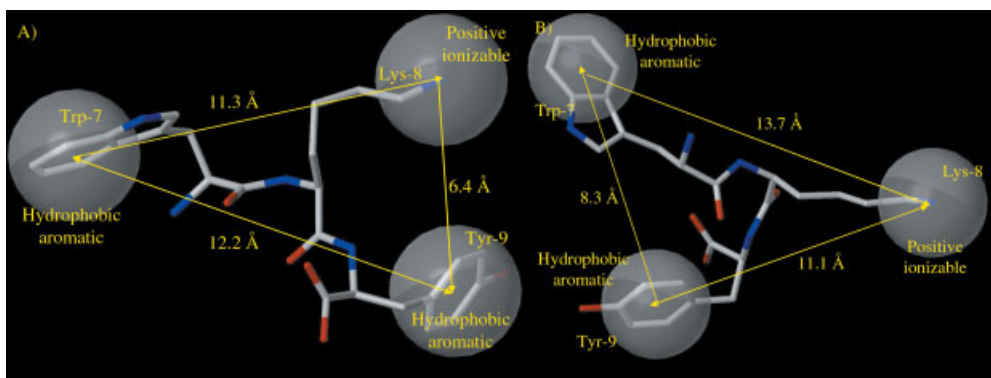
Fig. 9.6 The three-point pharmacophore models used for virtual screening obtained using the solution structures of (A) urotensin II and (B) the conformationally restricted hexapeptide Ac-c[CFwKYC]-NH$_2$. (After [50]).

hydrophobic aromatic centers from the tryptophan and tyrosine residues. This pharmacophore model is also shown in Fig. 9.6.

Screening of the Aventis compound collection identified 418 compounds matching the conformationally restricted hexapeptide-based pharmacophore, with only one compound exhibiting an IC$_{50}$ value below 10 μM. Using the urotensin II-based pharmacophore, 500 hits were found with 10 compounds, containing six different scaffolds, exhibiting IC$_{50}$ values between 400 nM and 7 μM. Nine of the 11 compounds identified obeyed the "rule of five" while the most active class of compounds, the substituted indoles, showed good metabolic stability in S9 cell lines but poor permeability as determined in experiments using Caco-2 cell monolayers. To improve the pharmacokinetic profile of these compounds, an optimization program was initiated to replace the benzamidine moiety with bioisosteres. A representative compound from the largest class, containing four active compounds, is shown in Fig. 9.7.
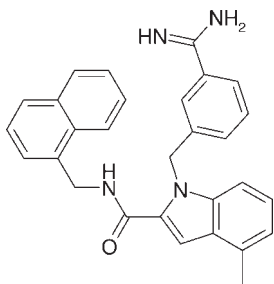
Fig. 9.7 The most potent compound, S6717, identified from virtual screening. S6717 was shown to have an IC$_{50}$ of 400 nM.

9.4.4
Melanin-concentrating Hormone-1 Receptor Antagonists

Obesity has been shown to be a major risk factor for a number of diseases including type-2 diabetes, stroke and certain cancers. Consequently, the pharmaceutical industry has responded by initiating a number of strategies primarily focusing on methods to reduce energy intake. Melanin-concentrating hormone, a 19-amino acid cyclic peptide, has been known to play a role in appetite stimulation for a number of years. However, it was only relatively recently that the orphan SLC-1 receptor was identified as the MCH receptor (MCH-1R). Experiments have revealed that stimulation of this receptor is correlated with an increase in food intake in rats. Conversely, continuous antagonism of the receptor results in a decrease in food intake coupled with a reduction in weight. Clark et al. [51] have used a number of virtual screening techniques to identify several different chemical series of MCH-1R antagonists.

The compound collections from a number of suppliers were collated and filtered to remove non-drug-like compounds, giving a total of approximately 615,000 compounds. A subset of this database was then interrogated using multiple virtual screening techniques. A total of 11 compounds, identified from the literature, were used as queries for 2-D substructure, 2-D similarity, 3-D similarity and 3-D substructure searches. A number of these compounds were also docked into a MCH-1R homology model and their docked conformations served as templates for pharmacophore searches. To identify additional compounds that may have been missed by previous searches, a clustering procedure was used. A selection of templates used during the virtual screening program is shown in Fig. 9.8.
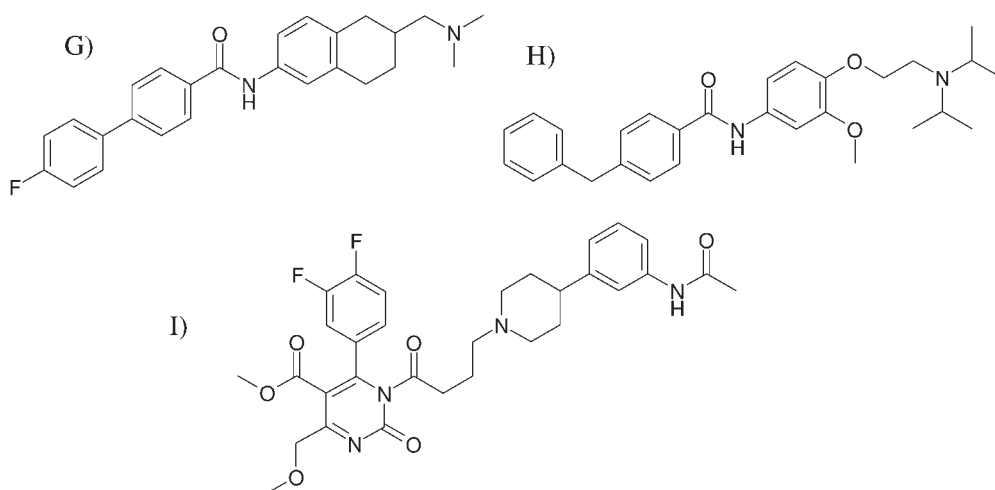
Fig. 9.8 A selection of the templates used during the virtual screening program:
(G) Takeda (T-226296), (H) GSK (SB568849) and (I) Synaptic (SNEC-3).

A total of 3015 compounds was identified from all the different searches. After removal of duplicates (i.e. compounds that were hits in multiple searches), the compounds were visually assessed in conjunction with a medicinal chemist. The purpose of this was to evaluate the compounds' synthetic tractability, novelty and drug-likeness together with their computed properties including molecular weight, polar surface area and ClogP. At the end of this exercise, 1490 unique compounds remained. To reduce this number further, the compounds were clustered using Ward's method based on Daylight fingerprints and a Tanimoto similarity threshold of 0.85, which resulted in 874 clusters. From these clusters, a total of 877 compounds were selected for purchase. After investigating availability and price, 806 compounds were ordered, 795 of which were received and screened.

Of the 795 compounds tested at a concentration of 10 µM, 62 active (> 40 % inhibition) compounds were identified and IC_{50} values were determined for 19 of these compounds. Four of these compounds demonstrated IC_{50} values below 1 µM, four had values between 1 and 10 µM and the IC_{50} values of the remaining 11 compounds lay in the range of 10–30 µM. The most potent compound identified, shown in Fig. 9.9, had an IC_{50} value of 55 nM with further testing confirming this compound to be a competitive antagonist.
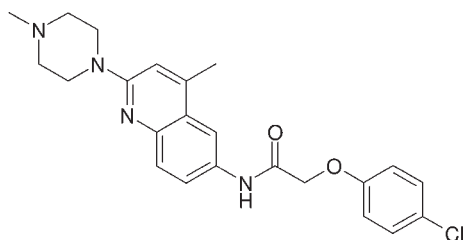
Fig. 9.9 Structure of the lead compound identified from the first round of virtual screening. This compound had an IC_{50} of 55 nM.

A retrospective analysis revealed that only two of the 11 compounds used as queries failed to yield a hit compound. In addition, nine of the 19 hit compounds were identified by one or more search methods. The 2-D similarity searches, with a hit rate of 5.6 %, appeared to be the most successful method employed, with the 3-D substructure searches (4.3 %) also performing well. Clustering and 3-D similarity searches yielded reasonable hit rates of 2.4 and 2.1 % respectively, while the 2-D substructure and site-directed pharmacophore failed to identify any hit compounds. This analysis confirmed the received wisdom that employing multiple search techniques with as many query compounds as possible helps to maximize the likelihood of identifying hit compounds [16].

9.4.5
Growth Hormone Secretagogue Receptor Agonists

The release of growth hormone from the pituitary gland is controlled by growth hormone secretagogue. Deficiency of growth hormone can occur in both the young and the elderly and also in people with severe burns, heart failure and AIDS. Adults suffering from growth hormone deficiency can experience decreased muscle mass, increased body fat, a lower bone density and a predisposition to atherosclerosis. Shoda et al. [52] have used two methods, similarity searching and cell-based compound partitioning, to identify new growth hormone secretagogue receptor agonists.

Two compounds, shown in Fig. 9.10, were identified from the patent literature and used to search a compound collection containing approximately 80,000 structures. For similarity searching, molecules were represented as "mini-fingerprints" and similarity was measured using the Tanimoto coefficient. The cell-based partitioning approach, on the other hand, divides chemical descriptor space into series of cells, or partitions, and then places compounds into their respective cells.
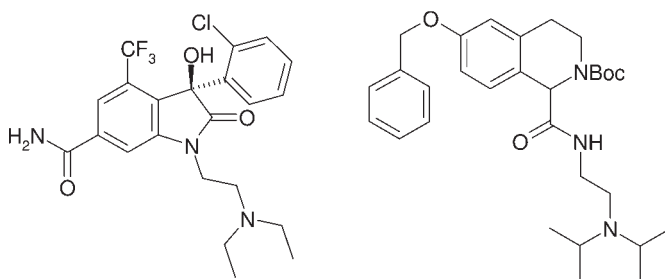
Fig. 9.10 The two compounds used as templates for virtual screening: SM-130686 (Sumitomo Pharmaceuticals, left) and a compound from WO-0185695 (Bristol-Myers Squibb).

About 100 top-scoring compounds were selected from similarity searching around both templates. After filtering, 58 compounds were selected with Tanimoto coefficients between 0.6 and 0.8, which is the preferred similarity range for the fingerprints used in this work. The cell-based partitioning searches identified 61 compounds that were placed in the same cells as the template molecules. Of these compounds 11 were removed because similarity searching had previously identified them, giving rise to a total of 108 compounds for testing.

Five compounds were found to be active after testing in a Ca^{2+} flux assay, equating to a hit rate of ~ 5%. Two of the compounds were identified from the similarity searches while the remaining three compounds came from the partitioning approach. The most potent compound identified by virtual screening, shown in Fig. 9.11, gave a relative agonistic activity of 57% (compared to a reference compound) at 10 µM and was selected for further SAR analysis.
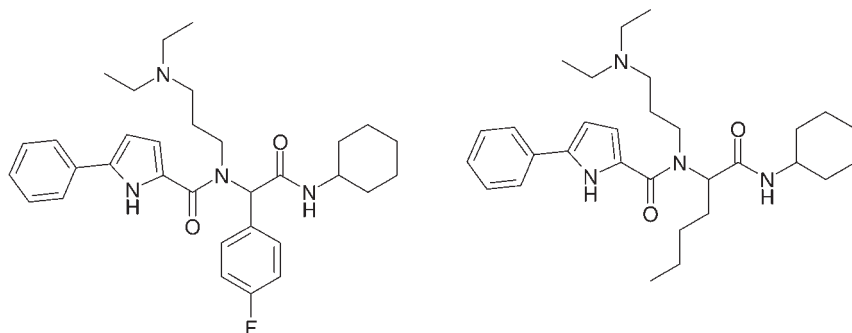
Fig. 9.11 The most potent compound identified by virtual screening (left) and an improved analog resulting from medicinal chemistry exploration.

Various compounds in which the diethylaminopropyl group was replaced were synthesized. Despite the similarity of some of the substituents used, including a morpholinopropyl group, a significant reduction in, or complete loss of, activity resulted suggesting that this group is critical for activity. Replacement of the 5-phenylpyrrolyl moiety also failed to improve the activity of these compounds. However, replacement of the 4-fluorophenyl by a number of groups, including 4-pyridyl and methyl, resulted in a significant increase in activity. The compound with a butyl group in this position (see Fig. 9.11) was found to be the most potent (160 % at 10 μM relative to the reference compound).

The 14 analogs generated during the SAR analysis that exhibited activity in the Ca^{2+} flux assay were subsequently tested in a radioligand binding assay to confirm direct binding to the receptor. Of these compounds, only one compound failed to show binding at a detectable level. The remaining compounds showed K_i values between 0.22 and 2 μM.

9.5
Conclusions

The examples cited above clearly demonstrate the effectiveness of ligand-based virtual screening as a hit-finding paradigm for GPCR targets. Given successes such as these, and the relative ease and rapidity with which such searches can be carried out, it is the authors' contention that ligand-based hit-finding approaches should be routinely applied in GPCR drug discovery projects whenever the available information permits. Looking ahead, it seems likely that the combination of ligand- and structure-based methods will allow researchers to benefit from the natural complementarity and synergy of the two approaches. Early reports of successes with such strategies are beginning to appear [10, 53, 54] promising a bright future for the discovery of new therapeutic agents acting at GPCRs.

References

1 NAMBI, P. and N. AIYAR. G protein-coupled receptors in drug discovery. *Assay Drug Dev Technol* **2003**; *1*: 305–310.

2 KLABUNDE, T. and G. HESSLER. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**; *3*: 928–944.

3 BLEICHER, K. H., L. G. GREEN, R. E. MARTIN, and M. ROGERS-EVANS. Ligand identification for G-protein-coupled receptors: a lead generation perspective. *Curr Opin Chem Biol* **2004**; *8*: 287–296.

4 PALCZEWSKI, K. et al. Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

5 BIKKER, J. A., S. TRUMPP-KALLMEYER, and C. HUMBLET. G-protein coupled receptors: models, mutagenesis, and drug design. *J Med Chem* **1998**; *41*: 2911–2927.

6 BALLESTEROS, J. A., L. SHI, and J. A. JAVITCH. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure–function analysis of rhodopsin-like receptors. *Mol Pharmacol* **2001**; *60*: 1–19.

7 BALLESTEROS, J. and K. PALCZEWSKI. G protein-coupled receptor drug discovery: implications from the crystal structure of rhodopsin. *Curr Opin Drug Discov Dev* **2001**; *4*: 561–574.

8 BISSANTZ C., P. BERNARD, M. HIBERT, and D. ROGNAN. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* **2003**; *50*: 5–25.

9 EVERS A. and G. KLEBE. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew Chem Int Ed Engl* **2004**; *43*: 248–251.

10 EVERS, A. and G. KLEBE. Successful virtual screening for a submicromolar antagonist of the neurokinin-1 receptor based on a ligand-supported homology model. *J Med Chem* **2004**; *47*: 5381–5392.

11 BECKER, O.M., S. SHACHAM, Y. MARANTZ, and S. NOIMAN. Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr Opin Drug Discov Dev* **2003**; *6*: 353–361.

12 ARCHER-LAHLOU, E. et al. Modeled structure of a G-protein coupled receptor: the cholecystokinin-1 receptor. *J Med Chem* **2005**; *48*: 180–191.

13 SHACHAM, S. et al. PREDICT modeling and in-silico screening for G-protein coupled receptors. *Proteins* **2004**; *57*: 51–86.

14 FURSE, K. E. and T. P. LYBRAND. Three-dimensional models for β-adrenergic receptor complexes with agonists and antagonists. *J Med Chem* **2003**; *46*: 4450–4462.

15 GOULDSON, P. R., N. J. KIDLEY, R. P. BYWATER, G. PSAROUDAKIS, H. D. BROOKS, C. DIAZ, D. SHIRE, and C. A. REYNOLDS. Toward the active conformations of rhodopsin and the β2-adrenergic receptor. *Proteins* **2004**; *56*: 67–84.

16 SHERIDAN, R. P. and S. K. KEARSLEY. Why do we need so many chemical similarity search methods? *Drug Discovery Today* **2002**; *7*: 903–911.

17 GREEN, D. V. Virtual screening of virtual libraries. *Prog Med Chem* **2003**; *41*: 61–97.

18 BAURIN, N. et al. Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million compounds. *J Chem Inf Comput Sci* **2004**; *44*: 643–651.

19 PEARLMAN, R. S. 3D molecular structures: generation and use in 3-D searching. In *3D QSAR in Drug Design: Theory, Methods and Applications*, Kubinyi, H. (Ed.). ESCOM: Leiden, **1993**.

20 SADOWSKI, J. and J. GASTEIGER. From atoms and bonds to three-dimensional atomic coordinates: automatic model builders. *Chem Rev* **1993**; *93*: 2567–2581.

21 CONCORD. Balducci, R., C. M. McGarity, A. Rusinko III, J. Skell, K. Smith, and R. S. Pearlman (University of Texas at Austin). Distributed by Tripos, Inc., 1699 S. Hanley Rd.,

Suite 303, St. Louis, MO 63144–2913, USA. http://www.tripos.com

22 Corina. Distributed by Molecular Networks GmbH – Computerchemie, Nägelsbachstraße 25, 91052 Erlangen, Germany. http://www.mol-net.de.

23 LYNE, P. D. Structure-based virtual screening: an overview. *Drug Discovery Today* **2002**; *7*: 1047–1055.

24 Unity. Developed and distributed by Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144–2913, USA. http://www.tripos.com

25 ISIS. Developed and distributed by Elsevier MDL, 14600 Catalina Street, San Leandro, CA 94577, USA. http://www.mdli.com

26 Catalyst. Developed and distributed by Accelrys, Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA. http://www.accelrys.com

27 FlexS. Developed and distributed by Bio-SolveIT GmbH, An der Ziegelei 75, 53757 Sankt Augustin, Germany. http://www.biosolveit.de. Also distributed by Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144–2913, USA. http://www.tripos.com

28 ROCS. Developed and distributed by OpenEye Scientific Software, 3600 Cerrillos Rd., Suite 1107, Santa Fe, NM 87507, USA. http://www.eyesopen.com

29 FeatureTrees. Developed and distributed by BioSolveIT GmbH, An der Ziegelei 75, 53757 Sankt Augustin, Germany. http://www.biosolveit.de

30 MOE. Developed and distributed by 1010 Sherbrooke St. West, Suite 910, Montreal, Quebec, H3A 2R7, Canada. http://www.chemcomp.com

31 Daylight Chemical Information Systems, 27401 Los Altos – Suite 360, Mission Viejo, CA 92691, USA. http://www.daylight.com

32 LEACH, A.R. and V. J. GILLET. *An Introduction to Chemoinformatics.* Kluwer Academic Publishers: Dordrecht, **2003**.

33 DOWNS, G.M and P. WILLETT. Similarity searching in databases of chemical structures. In *Reviews in Computational Chemistry*, vol. 7, Lipkowitz, K. B. and D. B. Boyd (Eds). VCH Publishers: New York, **1995**, 1–66.

34 MARTIN, Y. C., M. G. BURES, and P. WILLETT. Searching databases of three-dimensional structures. In *Reviews in Computational Chemistry*, vol. 1, Lipkowitz, K. B. and D. B. Boyd (Eds). VCH Publishers: New York, **1990**, 213–263.

35 GOOD, A. C. and J. S. MASON. Three-dimensional structure database searches. In *Reviews in Computational Chemistry*, vol. 7, Lipkowitz, K. B. and D. B. Boyd (Eds). VCH Publishers: New York, **1995**, 67–117.

36 PEPPERRELL, C. A. and P. WILLETT. Techniques for the calculation of three-dimensional structural similarity using inter-atomic distances. *J Comput-Aided Mol Des* **1991**; *5*: 455–74.

37 DRAYTON, S.K., K. EDWARDS, N. JEWELL, D. B. TURNER, D. J. WILD, P. WILLETT, P. M. WRIGHT, and K. SIMMONS. Similarity searching in files of three-dimensional chemical structures: Identification of bioactive molecules. *Internet J Chem* **1998**. http://www.ijc.com/ abstracts/abstract1n37.html.

38 MASON, J. S., A. C. GOOD, and E. J. MARTIN. 3-D pharmacophores in drug discovery. *Curr Pharm Des* **2001**; *7*: 567–597.

39 GRANT, J. A., M. A. GALLARD, and B. T. PICKUP. A fast method of molecular shape comparison: a simple application of a Gaussian description of molecular shape. *J Comput Chem* **1996**; *17*: 1653–1666.

40 HOFFMANN, R.D., S. MEDDEB, and T. LANGER. Use of 3D pharmacophore models in 3D database searching. In *Computational Medicinal Chemistry for Drug Discovery*, Bultinck, P. (Ed.). Marcel Dekker: New York, **2004**, 461–482.

41 YANG, L. et al. Synthesis and biological activities of potent peptidomimetics selective for somatostatin receptor subtype 2. *Proc Natl Acad Sci USA* **1998**; *95*: 10836–10841.

42 YANG, L. et al. Spiro1H-indene-1,4-piperidine. Derivatives as potent and selective non-peptide human somatostatin receptor subtype 2 (sst2) agonists. *J Med Chem* **1998**; *41*: 2175–2179.

43 CHARMm. Developed and distributed by Accelrys, Inc., 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA. http://www.accelrys.com

44 Miller, M.D., R.P. Sheridan, and S.K. Kearsley. SQ: A program for rapidly producing pharmacophorically relevant molecular superpositions. *J Med Chem* **1999**; *42*: 1505–1514.

45 Ariens, E.J., A.J. Beld, J.F. Rodrigues de Miranda, and A.M. Simonis. In *The Receptors: A Comprehensive Treatise*, O'Brien, R.D. (Ed.). Plenum: New York, **1979**, 33–91.

46 Marriott, D.P., I. G. Dougall, P. Meghani, Y.-J. Liu, and D. R. Flower. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M3 receptor antagonists. *J Med Chem* **1999**; *42*: 3210–3216.

47 Sybyl. Developed and distributed by Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144–2913, USA. http://www.tripos.com

48 Allen, F.H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr* **2002**; *B58*: 380–388.

49 DISCO (now DiscoTech). Developed and distributed by Tripos, Inc., 1699 S. Hanley Rd., Suite 303, St. Louis, MO 63144–2913, USA. http://www.tripos.com

50 Flohr, S., M. Kurz, E. Kostenis, A. Brkovich, A. Fournier, and T. Klabunde. Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure–activity relationships and nuclear magnetic resonance studies on urotensin II. *J Med Chem* **2002**; *45*: 1799–1805.

51 Clark, D.E., C. Higgs, S. P. Wren, H. J. Dyke, M. Wong, D. Norman, P. M. Lockey, and A. G. Roach. A virtual screening approach to finding novel and potent antagonists at the melanin-concentrating hormone 1 receptor. *J Med Chem* **2004**; *47*: 3962–3971.

52 Shoda, M. et al. Identification of structurally diverse growth hormone secretagogue agonists by virtual screening and structure–activity relationship analysis of 2-formylaminoacetamide derivatives. *J Med Chem* **2004**; *47*: 4286–4290.

53 Varady, J., X. Wu, X. Fang, J. Min, Z. Hu, B. Levant, and S. Wang. Molecular modeling of the three-dimensional structure of dopamine 3 (D3) subtype receptor: Discovery of novel and potent D3 ligands through a hybrid pharmacophore- and structure-based database searching approach. *J Med Chem* **2003**; *46*: 4377–4392.

54 Moro, S. et al. Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study. *J Med Chem* **2005**; *48*: 180–191.

10
3-D Structure of G Protein-coupled Receptors

Leonardo Pardo, Xavier Deupi, Cedric Govaerts, and Mercedes Campillo

10.1
Introduction

Membrane receptors coupled to guanine nucleotide-binding regulatory proteins (commonly known as G protein-coupled receptors, GPCRs) comprise one of the widest and most adaptable families of cellular sensors, as they are able to mediate a wide range of transmembrane signal transduction processes [1–4]. Recent estimations have assigned more than 1000 members to this family (http://www.ebi.ac.uk/integr8/ProteomeAnalysisAction.do?orgProteomeID=25). GPCRs are receptors for sensory signals of external origin such as odors, pheromones, or tastes; and for endogenous signals such as neurotransmitters, (neuro)peptides, divalent cations, proteases, glycoprotein hormones, purine ligands, and others [4]. GPCRs constitute one of the most important pharmaceutical targets, as around 40% of clinically prescribed drugs act through these receptors [5, 6]. Among the 100 top-selling drugs 25% are targeted at members of this protein family and their annual worldwide sales exceeded $30 billion in 2001 [7].

The biological function of GPCRs is commonly discussed in terms of the extended ternary complex model, which proposes an equilibrium between inactive and active states [8, 9]. In this model agonists bind the active form of the receptor with higher affinity than the inactive form, thus shifting the equilibrium towards the active state of the receptor; inverse agonists have higher affinity for the inactive form, decreasing the population of the active state; and neutral competitive antagonists have equal affinity for both forms of the receptor. GPCRs are kept in the inactive conformation through a network of constraining intramolecular interactions [3, 10–12]. Release of these interactions, by ligand binding, by the interaction with the G protein, or by the so-called constitutively active mutations (CAM, mutant receptors that present a significant increase in basal signal in the absence of agonist [13]), leads to the co-existence of different active conformations of the receptor [14, 15].

As a result of the low natural abundance of GPCRs and the difficulty in producing and purifying recombinant protein, only one member of this family, bovine
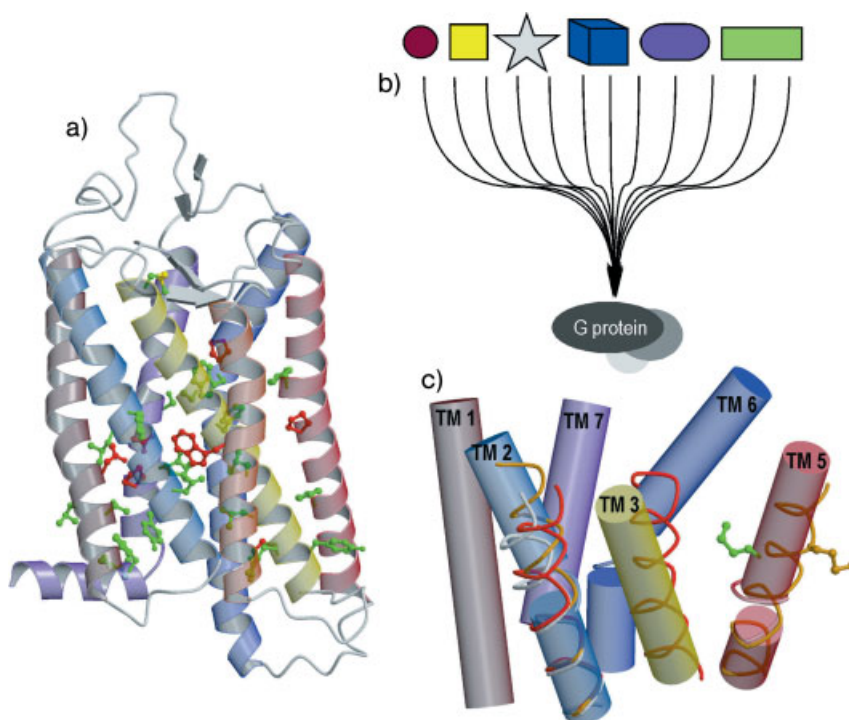
Fig. 10.1 (a) Crystal structure of bovine rhodopsin (PDB code 1GZM). The highly conserved amino acids in the rhodopsin family of GPCRs in each TM helix N1.50 (100%), D2.50 (94%), R3.50 (96%), W4.50 (96%), P5.50 (77%), P6.50 (100%), and P7.50 (96%) are shown in red (standardized nomenclature [32]). Amino acids that are conserved in more than 80% of the receptor sequences [31] are shown in green. (b) The sequence of the central part and cytoplasmic ends of the TM helices are more conserved than the extracellular domains. Thus, despite the structurally diverse type of extracellular signals (circle, square, star, cube, oval, rectangle) the processes that propagate the signal from the ligand binding site to the intracellular amino acids of the trans-membrane bundle must converge towards structurally conserved mechanisms of effector activation. While the first steps of these mechanisms will be specific for each subfamily, the final steps will share many common structural features. (c) The transmembrane helices of rhodopsin (except TM4) are shown as cylinders. Computer simulated TM helices (tube ribbon) containing polyAla (control, white), Pro at position 2.59 (gold), and the T2.56xP2.58 motif (red) superimposed on the cytoplasmic end of TM2. The bending of TM2 toward TM3 results in the relocation of TM3 (shown in red) toward TM5. A regular Pro-kink (gold) is superimposed on the intracellular domain of TM5. These figures were created using MolScript v2.1.1 [83] and Raster3D v2.5 [84].

rhodopsin, the light photoreceptor protein of rod cells, has been crystallized so far [16–19]. Five structural models of rhodopsin are available at the Protein Data Bank, at resolutions of 2.8 Å (PDB identifiers 1F88 and 1HZX), 2.65 Å (1GZM), 2.6 Å (1L9H), and 2.2 Å (1U19). Rhodopsin is formed by an extracellular N-terminus of four β-strands, seven transmembrane helices (TM1 to TM7, which has also led to this family of proteins being known as 7TM receptors) connected by alter-

nating intracellular (i1 to i3) and extracellular (e1 to e3) hydrophilic loops, a disulfur bridge between e2 and TM5, and a cytoplasmic C-terminus containing an α-helix (HX 8) parallel to the cell membrane (see Fig. 10.1a). Several reviews on the structure of rhodopsin have been published [20, 21].

The availability of the rhodopsin structure allows the use of homology modeling techniques for building three-dimensional models of other homologous GPCRs [22]. Here, we evaluate both the three-dimensional structural models and the molecular mechanism for agonist-induced activation of rhodopsin-like G protein-coupled receptors.

10.2
Classification of G Protein-coupled Receptors

The analysis of the human genome has led to the classification of the GPCR sequences into five main families: rhodopsin, glutamate, adhesion, frizzled/taste2, and secretin [23]. The rhodopsin family is the largest and is subdivided into four main groups (α, β, γ, δ) with 13 sub-branches (α, prostaglandin, amine, opsin, melatonin, MECA; β, peptides; γ, SOG, MCH, chemokine; δ, MAS, glycoprotein, purine, olfactory). These groups include orphan GPCRs, receptors for which the ligand and the (patho)physiological function remain unknown. Pairing orphan GPCRs to potential ligands found in tissue extracts or in libraries is an area of active research [24]. Phylogenetic and receptor–ligand relationships among 367 human receptors for endogenous ligands showed 143 orphan receptors in 2003 [25]. However, this number decreases at the rate of seven to eight deorphanized receptors per year [24].

10.3
The Extracellular N-terminal Domain of G Protein-coupled Receptors

The structures of the N-terminal extracellular segment of several GPCRs, in addition to rhodopsin, have been elucidated: the glutamate receptor (PDB identifier 1EWK) [26]; the Drosophila mutant Methuselah, related to the secretin family of GPCRs (1FJR) [27]; the frizzled receptor (1IJY) [28]; the corticotropin-releasing factor (1U34) [29]; and the follicle stimulating hormone receptor in complex with its hormone (1XWD) [30]. These receptors differ greatly at the sequence and structure levels, which support the idea of divergence in this region among GPCRs.

10.4
Sequence Analyses of the 7TM Segments of the Rhodopsin Family
of G Protein-coupled Receptors

Statistical analysis of the residues forming the TM helices of the rhodopsin family of GPCRs shows a large number of conserved sequence patterns [31]. Importantly

at least one amino acid is remarkably conserved in each TM: N in TM1 (100%), D in TM2 (94%), R in TM3 (96%), W in TM4 (96%), P in TM5 (77%), P in TM6 (100%), and P in TM7 (96%). These amino acids are easily identifiable on a multiple sequence alignment and serve as reference points to define a general numbering scheme that applies to all rhodopsin-like GPCRs and allows easy comparison between residues in the 7TM segments of different receptors [32]. Each residue is identified by two numbers: the first (1 through 7) corresponds to the helix in which it is located; the second indicates its position relative to the most conserved residue in the helix, arbitrarily assigned to 50. For instance, N7.49 is the asparagine in TM7, located one residue before the highly conserved proline P7.50. Figure 10.1a shows the position, in the crystal structure of rhodopsin, of these amino acids (red) and other amino acids that are conserved in more than 80% (green) of receptor sequences. Clearly, the large number of conserved sequence patterns in these 7 TM segments suggests that the helical bundle of rhodopsin is common to the GPCR family. This structural homology between rhodopsin and the other GPCRS cannot be extended to the extracellular domain, for which there is very little sequence conservation. In rhodopsin, it is highly structured, blocking the access of the solvent to the core of the receptor and retinal, the covalently bound ligand, a scenario unlikely to occur among receptors responding to diffusible ligands [33].

10.5
The Conformation of Pro-kinked Transmembrane α-Helices

Although Pro has the least helix-forming tendency [34] and the highest turn-stabilizing tendency in the membrane [35] of all naturally occurring amino acids, Pro residues are normally observed in TM helices [36] where they usually induce a significant distortion named Pro-kink [37]. The break is produced in order to avoid a steric clash between the pyrrolidine ring of Pro (at position i) and the carbonyl oxygen of the residue in the preceding turn (position i – 4) [38] which produces a local opening of the α-helical structure, leading to a bending of approximately 20° (Fig. 10.2) [39, 40]. In addition, the absence of the backbone N-H group in Pro prevents the formation of the hydrogen bond with the backbone carbonyl oxygen of residue i – 4. Therefore, in the context of an α-helix, Pro not only induces a break in the helix, but it also destabilizes the hydrogen bond network that normally stabilizes the secondary structure. As a result, Pro introduces a flexible point in the α-helix which could be of functional importance for the membrane protein [41]. It has thus been suggested that the highly conserved Pro residues in TMs 5, 6, and 7 in the rhodopsin family of GPCRs are involved in mechanisms of signal transduction [3, 42].

Several procedures have been described for measuring the distortion of transmembrane α-helices [43, 44]. In this chapter we will refer to two relevant parameters. First, the bend angle, defined as the angle between the axis of the cylinders formed by the residues preceding and following the motif that induces the
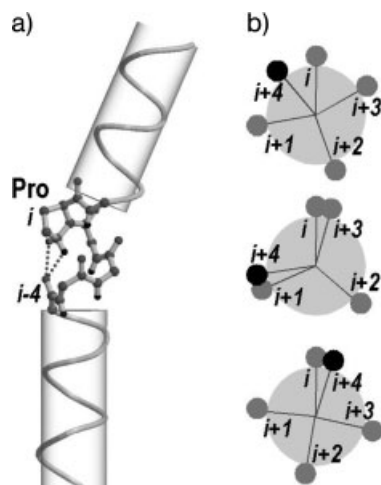
Fig. 10.2 (a) A detailed view of the hydrogen bond pattern in the Pro-kink turn. (b) Position of residues i, i + 1, i + 2, i + 3, and i + 4 in an ideal α-helix (top panel) of approximately 3.6 residues per turn (twist =100°), a closed helix (middle) of <3.6 residues per turn (twist >100°), and an open helix (bottom) of >3.6 residues per turn (twist <100°).

distortion in the helix (Fig. 10.2 a). Second, a residue-residue twist angle, or unit twist, calculated for sets of four consecutive Cα atoms, i.e. one turn, to analyze helical uniformity [44]. An ideal α-helix (Fig. 10.2 b, top panel), with approximately 3.6 residues per turn, has a unit twist of approximately 100° (360°/3.6). A closed helical segment, with <3.6 residues per turn, possesses a unit twist of >100° (middle panel), whereas an open helical segment, with >3.6 residues per turn, possesses a unit twist of <100° (bottom panel). A variation greater than 20° in the unit twist angle will result in a change in the orientation of the amino acid side chain and, in addition, local opening or closing will result in a change of the structure of the helix, making it to point towards different positions in space.

10.6
Helix Deformation in the Rhodopsin Family of G Protein-coupled Receptors

The TM helices in rhodopsin are not regular α-helices [19, 21]. On the contrary, these helices are highly bent and distorted to facilitate packing in the context of the membrane. Here we present a detailed analysis of these distortions and their implication for modeling other GPCRs [45].

10.6.1
Transmembrane Helix 1

Figure 10.3 b shows the conformation of a poly-Ala ideal α-helix (control, white), a standard Pro-kinked α-helix (gold), and TM1 of rhodopsin (dark red). Clearly, TM1 of rhodopsin is halfway between the ideal poly-Ala and Pro-kinked α-helices. Despite the presence of P1.48 in the amino acid sequence of rhodopsin, TM1 has a bend angle of only 9° [19], much lower than the average bend angle of 20° of
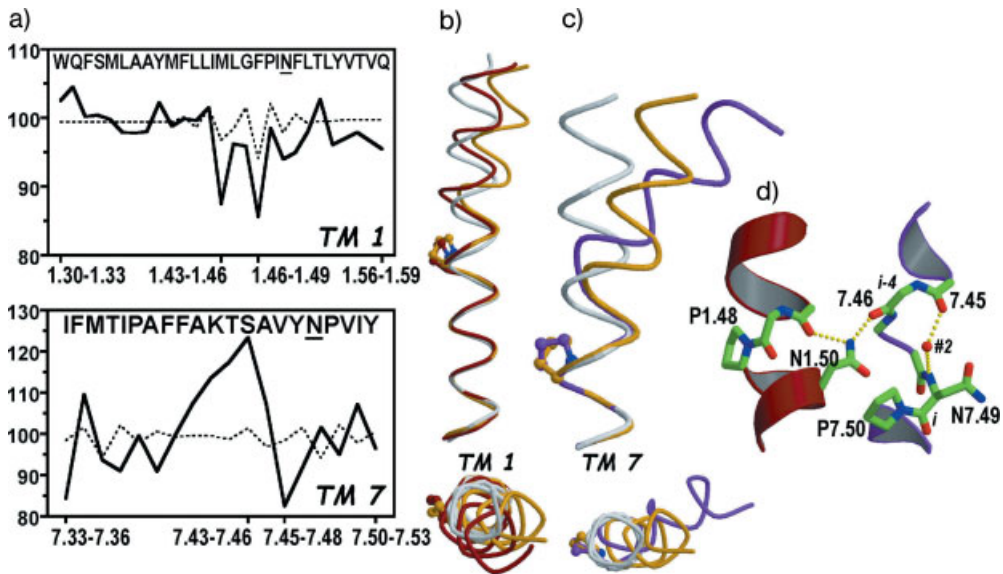
Fig. 10.3 (a) Evolution of the unit twist angles (°) of a standard Pro-kink helix (dashed line) [40] and TMs 1 (top) and 7 (bottom) of rhodopsin (solid line). The amino acid sequence of rhodopsin is shown (N1.50 and P7.50 are underlined). (b, c) Comparison of helix bending between an ideal α-helix (control, white), a standard Pro-kink (gold), and TMs 1 (panel b, dark red) and 7 (panel c, purple) of rhodopsin (different views rotated by 90°). The side chain of Pro is shown. (d) Detailed view of the interface between TMs 1 (dark red) and 7 (purple).

Pro-kinked α-helices [40]. Figure 10.3 a (top panel) shows the unit twist angle of a standard Pro-kinked α-helix (dashed line) [40] and TM1 of rhodopsin (solid line). TM1 of rhodopsin is open (>3.6 residues/turn, twist <100°) at the 1.43–1.46 and 1.46–1.49 turns. The highly conserved N1.50 most probably influences, in rhodopsin and other GPCRs, the conformation of TM1 (Fig. 10.3 d). The $N_{\delta2}$-H_2 atoms of N1.50 act as a hydrogen bond donor in the interaction with the carbonyl oxygens of residues at positions 1.46 and 7.46, linking TMs 1 and 7. The absence of P1.48 in many GPCRs (P1.48 is only present in 9% of the sequences [31]) would tend to decrease the bend of the helix. However, the effect is not expected to be significant because TM1 of rhodopsin already possesses a small Pro-kink-induced angle of 9°.

10.6.2
Transmembrane Helix 2

TM2 provides a remarkable example of local structural variability in the GPCR family. With a limited sequence-variability, subfamilies of receptors have evolved different structural motifs. Rhodopsin contains two non-conserved successive Gly residues at positions 2.56 and 2.57. This specific motif of the opsin family induces
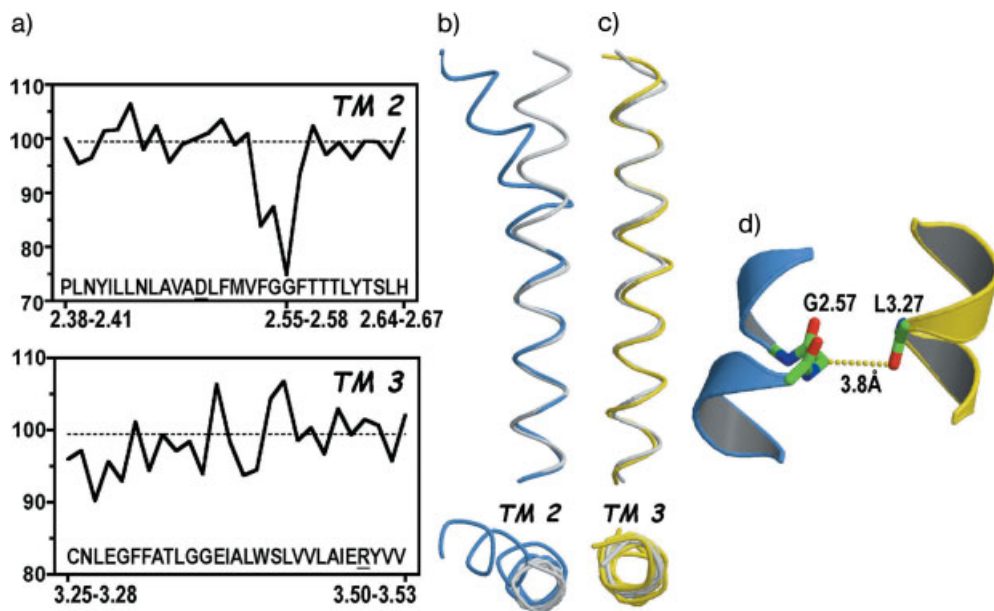
Fig. 10.4 (a) Evolution of the unit twist angles (°) of an ideal α-helix (dashed line) and TMs 2 (top) and 3 (bottom) of rhodopsin (solid line). The amino acid sequence of rhodopsin is shown (D2.50 and R3.50 are underlined). (b, c) Comparison of helix bending between an ideal α-helix (control, white) and TMs 2 (panel b, light blue) and 3 (panel c, yellow) of rhodopsin (different views rotated by 90°). (d) Detailed view of the interface between TMs 2 (light blue) and 3 (yellow).

a significant distortion of TM2, which bends 33° [19] towards TM 1 (Figs. 10.1 and 10.4) and opens (>3.6 residues/turn, twist <100°) at the 2.55–2.58 turn (Fig. 10.4a, top panel). The C_α-H group of G2.57 forms an interhelical hydrogen bond with the backbone carbonyl of the residue at position 3.27 (Fig. 10.4d). This type of C_α-H···O=C hydrogen bond is an important determinant of stability and specificity in membrane protein folding [46].

Most of the other members of the rhodopsin family of GPCRs possess at this 2.57 position a bulky β-branched or γ-branched amino acid (Leu, 39%; Val, 14%; Ile, 3%) that prevents the interaction with the carbonyl of residue 3.27 as seen in rhodopsin. It seems reasonable to suggest that these bulky side chains would clash with TM3, relocating its position (see below). Furthermore, the angiotensin, opioid, and chemokine family of GPCRs possess in this region of TM2 a conserved T2.56XP2.58 motif (X being a non-conserved residue). P2.58 orients the extracellular end of TM2 toward TM3 and away from TM1 (Fig. 10.1c). Moreover, the presence of Thr in the TxP motif increases the helix bend angle by about 7–10° [40], causing the extracellular side of TM2 to lean even more toward TM3 and slightly toward the center of the bundle [47]. In contrast, biogenic amine receptors contain a Pro residue at position 2.59. This proline would bend the helix in the direction of TM1 (Fig. 10.1c). Figure 10.1c shows representative structures

of helices containing polyAla (control, white), Pro at position 2.59 (gold), and the T2.56xP2.58 motif (red) superimposed on the cytoplasmic end of TM2 of the rhodopsin structure. Clearly, the different amino acid sequences that form TM2 in rhodopsin and other homologous GPCRs produce structural divergences in this region of TM2 and the nearby TM3 (see below).

10.6.3
Transmembrane Helix 3

Figure 10.4 a (bottom panel) shows the unit twist angle of TM3 of rhodopsin, and Fig. 10.4 c shows the conformation of a poly-Ala α-helix (control, white) and TM3 of rhodopsin (yellow). Both methods reveal no major structural distortions in TM3 of rhodopsin [19]. However, as detailed above, sequence divergence between rhodopsin and the other GPCRs indicates that for most receptors, TM2 should bend toward TM3, resulting in the subsequent relocation of TM3 toward TM5 (Fig. 10.1 c) [48–50]. The relocation of TM3 in some GPCRs is aided by the presence of several Ser or Thr residues in the sequence. The formation of a second hydrogen bond between the side chain of Ser or Thr and the peptide carbonyl oxygen in the previous turn of the helix induces or stabilizes moderate but significant local bending [51]. Incorporation of this small bending angle in one end of the TM helix results in a significant displacement of the residues located at the other end. We have suggested that Thr3.37, 85% conserved in the neurotransmitter family of GPCRs, produces such effect [48]. The relocation of the central TM3 accounts for interactions between the neurotransmitters and Asp3.32 in TM3 and a series of conserved Ser/Thr residues (5.42, 5.43 and 5.46) in TM5 that were demonstrated experimentally [52].

10.6.4
Transmembrane Helix 4

In the structure of bovine rhodopsin, TM 4 has a bend angle of 35 ° [19], higher than the average Pro-kink-induced angle of 20°, due to the presence of two consecutive Pro residues at positions 4.59 and 4.60. This Pro-Pro motif is not conserved and is only present in some vertebrate opsins and some amine receptor subfamilies. In contrast, other GPCRs possess a single Pro or the Pro-X-Pro motif at these positions. This lack of conservation suggests structural divergences at the extracellular part of TM4 and may be related to the structural requirements of the second extracellular loop.

10.6.5
Transmembrane Helix 5

Rhodopsin possesses the highly conserved P5.50 in TM5. However, TM5 has a bend angle of only 13° [19], much lower than the average Pro-kink angle (see Fig. 10.5 b and c). Structural analysis reveals a peculiar structural motif in rhodop-

sin where the Pro-kink-induced bend of TM5 is decreased due to an opening of the helix (> 3.6 residues/turn, twist < 100°) at the 5.45–5.48 turn (Fig. 10.5 a and d) and, thus, partially removing the steric clash between the pyrrolidine ring of P5.50 and the carbonyl oxygen at position 5.46 (Fig. 10.5 e). This carbonyl oxygen adopts an unusual and probably unstable conformation that is stabilized by the interactions with E3.37 (present in only 4 % of the sequences) and L3.40 (Fig. 10.5 e). The conservation pattern at position 3.40 suggests a functional role. The rhodopsin family of GPCRs possesses a bulky hydrophobic amino acid at this position in 76 % of the sequences (L, 9 %; V, 25 %; I, 42 %). This hydrophobic side chain probably forms a C–H···O=C interaction that stabilizes the carbonyl conformation (Fig. 10.5 e).
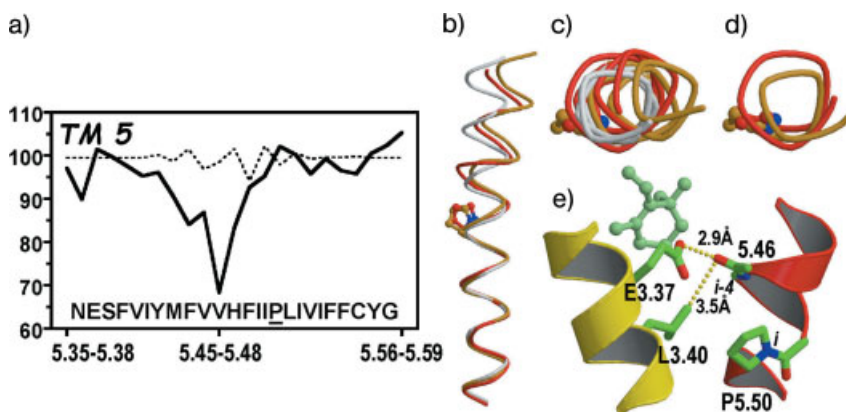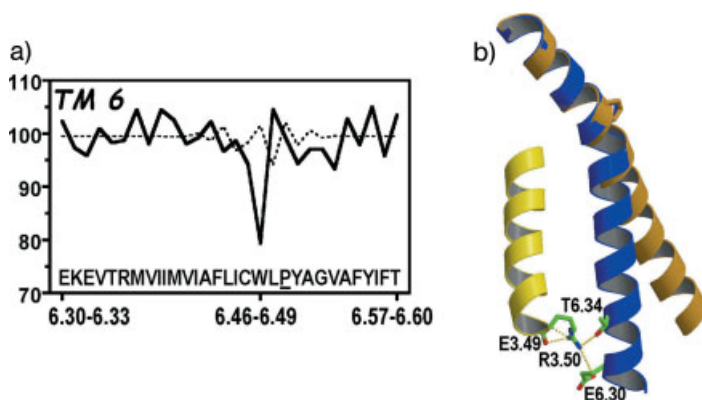

Fig. 10.5 (a) Evolution of the unit twist angles (°) of a standard Pro-kink helix (dashed line) and TM5 of rhodopsin (solid line). The amino acid sequence of rhodopsin is shown (P5.50 is underlined). (b, c) Comparison of helix bending between an ideal α-helix (control, white), a standard Pro-kink (gold), and TM5 (red) of rhodopsin (different views rotated by 90°). The side chain of Pro is shown. (d) Comparison of the opening of TM5 (red) relative to a standard Pro-kink helix (gold) at the 5.45–5.48 turn. (e) Detailed view of the interface between TMs 3 (yellow) and 5 (red).

The question arises whether TM5 of other Pro-containing GPCRs would adopt a conformation similar to rhodopsin or a more standard Pro-kink conformation. The opening of the helix at the 5.45–5.48 turn not only modifies the bending of the helix but also modifies the relative orientation of the side chains on either side of the Pro-kink, including residues involved in binding the extracellular ligand such as 5.42. For instance, R5.42 of the N-formyl, C3 a, and C5 a peptide receptors is involved in the interaction with the C-terminus carboxylate group of the peptide [53, 54] and S5.42 of the neurotransmitter family hydrogen bonds the hydroxyl group of many neurotransmitters [55]. Figure 10.1 c shows the orientation of the side chain at position 5.42 in the crystal structure of rhodopsin (green) and in a regular Pro-kink (gold). Clearly, the absence of the opening of the helix incorrectly orients the side chain towards the lipid environment. In contrast, opening of the

helix properly positions the side chain towards the binding site crevice. This data suggests a common conformation of TM5 in the rhodopsin family of GPCRs.

10.6.6
Transmembrane Helix 6

TM6 of rhodopsin (blue) possesses a bend angle of 35° [19] much higher than the average Pro-kink-induced (gold) angle as clearly shown in Fig. 10.6b. This extreme conformation of a Pro-containing helix is stable due to the ionic interaction between Glu6.30 and Arg3.50 of the (D/E)RY motif in TM3 [56]. To achieve this interaction TM6 must be opened (>3.6 residues/turn, twist <100°) at the 6.46–6.49 turn (Fig. 10.6a). Disruption of this ionic lock [56], aided by the protonation of (D/E3.49) [57], induces large conformational changes of TMs 3 and 6 [58], considered to be an essential step in the process of GPCR activation (see below).

Fig. 10.6 (a) Evolution of the unit twist angles (°) of a standard Pro-kink helix (dashed line) and TM6 of rhodopsin (solid line). The amino acid sequence of rhodopsin is shown (P6.50 is underlined). (b) Detailed view of the ionic lock linking the intracellular part of TMs 3 (yellow) and 6 (blue). A standard Pro-kink helix (gold) is superimposed on the extracellular part of TM6.

The acidic residue at position 6.30 is only conserved in 32% of sequences (D, 7%; E, 25%). Opioid receptors feature a Leu in 6.30, so the role of E6.30 is likely to be played by T6.34, through a similar although specialized set of intramolecular interactions with R3.50 [59]. However, many GPCRs contain a basic residue at this 6.30 position (34%; K, 18%; R, 16%) preventing direct interaction with R3.50. These receptors will probably possess a totally different network of interhelical interactions at the intracellular side that remains to be identified.

10.6.7
Transmembrane Helix 7

The TM7sequence in rhodopsin contains P7.38 and P7.50 of the highly conserved NPxxY motif. Figure 10.3 c shows the conformation of a poly-Ala ideal helix (control, white), a standard Pro-kinked α-helix (gold), and TM7 of rhodopsin (purple). TM7 of rhodopsin is highly kinked with estimated bend angles of 27° and 14° for the N- and C-terminal kinks, respectively [19]. Moreover, the helix is open (> 3.6 residues/turn, twist < 100 °) at the 7.45–7.48 turn and closed (< 3.6 residues/turn, twist > 100°) at the 7.43–7.46 turn (Fig. 10.3 a, bottom panel). Several structural factors distort TM7: first, the hydrogen bond between N1.50 and the backbone carbonyl oxygen at position 7.46 (Fig. 10.3 d). Opening of the helix at the 7.43–7.46 turn removes the steric clash between the pyrrolidine ring of P7.50 and the carbonyl oxygen at position 7.46 in a similar manner to that observed in TM5 (compare Fig. 10.3 d and e). The polar $N_{\delta2}$-H_2···O=C hydrogen bond interaction stabilizes the carbonyl conformation. Second, several structural water molecules have been found in the most recent crystal structures of rhodopsin [18, 19]. Water#2 is located between the backbone carbonyl at position 7.45 and the backbone N-H amide of N7.49 (Fig. 10.3 d). As a result of this additional interaction, the standard hydrogen bond distance of 3.0 Å between N_i and O_{i-4} is increased to 4.9 Å in the crystal structure of rhodopsin. Water#2 links two highly conserved side chains, D2.50 (94% of the sequences) and N7.49 (75%) [18, 19], thus, suggesting its presence in other GPCRs. Third, rhodopsin contains the Thr7.36xP7.38 at the top of TM7, exhibiting a larger bending angle (27° [19]) than a regular Pro-kinked helix (20° [39]) that matches the computationally determined bend of TxP (27° [40]). The absence of both Thr and Pro residues at the top of TM7 in other GPCRs suggests structural divergences in this region of TM7.

TM7 starts at position 7.33 (inclusive), as observed in the three-dimensional structure of rhodopsin. However, studies using the substituted cysteine accessibility method (SCAM), which maps the residues of TMs that are on the water-accessible surface of the binding-site crevices [60], show structural differences among GPCRs. MTSEA-sensitive mutants start at positions 7.28 in the AT1 receptor [61], 7.31 in the δ opioid receptor [62], or 7.34 in the κ opioid [62], D2 dopamine [63], and A1 adenosine [64] receptors. Thus, the TM7 may start at different positions in different GPCRs and have a dissimilar number of helical turns in the extracellular portion.

10.7
Structural and Functional Role of Internal Water Molecules

The most recent crystal structures of rhodopsin propose that water molecules in the vicinity of highly conserved residues regulate the activity of rhodopsin-like GPCRs, whereas water molecules in the retinal pocket regulate spectral tuning in visual pigments [18, 19]. The water molecules in the environment of highly con-

served residues are most likely present in all members of the family and they should be taken into account when modeling GPCRs.

10.7.1
A Conserved Hydrogen Bond Network Linking D2.50 and W6.48

Figure 10.7 a shows a detailed picture of the D2.50 and W6.48 environment of rhodopsin [19]. Water#10 and water#12 form a hydrogen bond network linking D2.50 and W6.48. Interestingly, a conformational transition of W6.48 upon receptor activation has been observed in the structure of metarhodopsin I [65]. Thus, the D2.50–water#12–water#10–W6.48 network is suggested to restrain the receptor in the inactive conformation (see below).

Of the sequences among the rhodopsin family of GPCRs 67% contain N7.45 [31]. N7.45 would be located in the same position as in rhodopsin with water#10
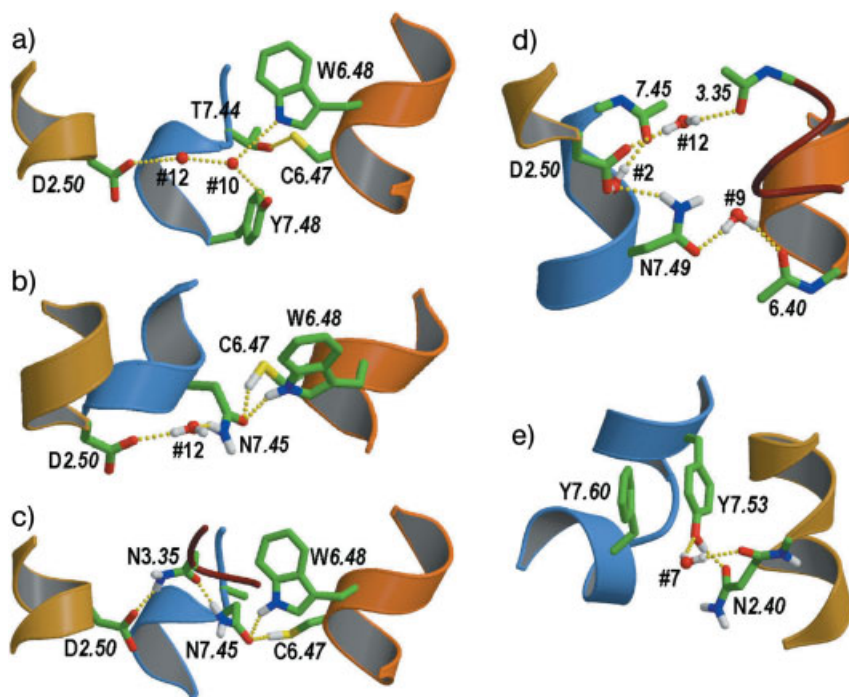
Fig. 10.7 (a–c) The proposed hydrogen bond network linking D2.50 and W6.48 in the inactive conformation of (a) rhodopsin (D2.50–water#12–water#10–W6.48) [19], (b) N7.45-containing receptors (D2.50–water#12–N7.45–W6.48) [66], and (c) N3.35/N7.45-containing receptors (D2.50–N3.35–N7.45–W6.48) [62]. (d) A detailed view of the proposed hydrogen bond network at the D2.50/N7.49 environment in the inactive conformation. (e) The interactions of Y7.53 of the NPxxY motif with the residues located in TM2 and HX8. The C_α traces of TMs 2 (golden red), 3 (dark red), 6 (orange), and 7 (blue) are shown.

linking W6.48 and water #12 (Fig. 10.7 b). Thus, N7.45-containing receptors are able to form a D2.50–water#12–N7.45–W6.48 network to maintain the receptor in the inactive state [66]. Similarly, opioid, bradykinin, formyl-met-leu-phe, somatostatin, angiotensin, C5 a anaphylatoxin, and proteinase-activated families of GPCRs possess N3.35 (29% of the receptors) in their sequences. N3.35 would mimic the role of water#12 (Fig. 10.7 c). Thus, N3.35/N7.45-containing receptors form the D2.50–N3.35–N7.45–W6.48 network of interactions [62]. Therefore, the number of water molecules linking D2.50 and W6.48 will depend on the type of side chain at positions 3.35 and 7.45.

10.7.2
The Environment of the NPxxY Motif in TM7

Figure 10.7 d shows a detailed view of waters#2, #9 and #12 found in the D2.50/ N7.49 environment of rhodopsin [19]. These structural waters mediate a number of interhelical interactions through the highly conserved D2.50 and N7.49 amino acids. The lack of coordination of the polar D2.50 and N7.49 in both rhodopsin and other GPCRs leads us to suggest that this set of water molecules is conserved in the family and is putatively involved in maintaining the inactive form of the receptor (see below). Water#2 is located between the 7.45 and 7.49 turn as described above (Fig. 10.3 b). Water#9 mediates interhelical interactions between the side chain of N7.49 and the backbone carbonyl at position 6.40, keeping the Asn side chain towards TM6 in the inactive state of the receptor [18, 67]. Water#12 links D2.50 with both the backbone carbonyl of the residue at position 3.35 (Fig. 10.7 d) and either water#10 (Fig. 10.7 a) or N7.45 (Fig. 10.7 b). Figure 10.7 e shows the interactions of Y7.53 with the residues located in TM2 and HX8, the cytoplasmic extension of TM7 which runs parallel to the membrane. The aromatic moiety of Y7.53 interacts with F7.60 in HX8 whereas the hydroxyl group of Y7.53 forms hydrogen bonds with the $N_{\delta 2}$-H_2 group and the carbonyl oxygen (via water#7) of the partly conserved N2.40 in TM2 (40% of the sequences among family A of GPCRs). The NPxxY motif is highly conserved in the rhodopsin family of GPCRs and plays a key role in the process of receptor activation by achieving key interhelical interactions (see below). These interactions, mediated through waters molecules, are presumably conserved in the family.

10.8
Molecular Processes of Receptor Activation

Conceptually, the processes of transmembrane signal transduction can be divided into two groups. First, those processes initiated by the recognition of the extracellular ligand by the receptor. These processes will extensively depend on the specific subtype of receptor since wide ranges of extracellular ligands, from small neurotransmitters to large hormones, are recognized by GPCRs. Each subfamily has probably developed specific structural motifs that allow the receptor to accommo-

date and respond to its cognate ligand (see Fig. 10.1 b); and second, the processes that propagate the signal from the ligand binding site to the amino acids of the cytoplasmic side of the transmembrane bundle. Despite the structurally diverse type of extracellular ligands, the sequence conservation pattern of GPCRs suggests that these processes might occur by means of common mechanisms. Specifically, there is a higher degree of sequence conservation in the central part and cytoplasmic ends of the TM helices compared to the extracellular domains [31]. As a result, the most conserved residues are clustered in the middle or in intracellular regions of the receptor (see Fig. 10.1 a). Figure 10.1 b illustrates this concept, the different lines representing converging structural processes as the signal progresses towards the intracellular part of the receptor. This convergence can be extended further as thousands of extracellular signals are recognized by hundreds of GPCRs, to activate dozens of G-proteins to become phosphorylated by seven G-protein receptor kinases, and to be capped by three arrestins [31, 68].

10.8.1
Molecular Processes Initiated by the Recognition of the Extracellular Ligand by the Receptor

Given the large number and diverse types of both ligands and receptors we will focus here on one detailed example to illustrate the process of receptor activation upon ligand binding: the β_2 adrenergic receptor. This typical family A GPCR can be obtained in considerable quantities for biophysical studies [69]. A kinetic analysis of the process of activation has shown that β_2 adrenergic receptor activation by agonists is a multi-step process [15, 70] (see Fig. 1 in [70] for an excellent illustration of the chemical structures of the ligands and their interaction with the receptor). Although the details of these steps have not been dissected yet, they can be hypothesized as follows: (1) formation of the ionic interaction between the highly conserved D3.32 in TM3 and the protonated amine of the ligand, and the aromatic–aromatic interaction between F6.52 in TM6 and the catechol ring of the ligand [71]; (2) relocation of TM5 so that S5.42, S5.43, and S5.46 can interact optimally with the catechol hydroxyl groups of the ligand [55, 72]; and (3) relocation of TM6 to stabilize the interaction between W6.48 (residues in the toggle switch [73], see below) and the ligand, and between N6.55 and the β-hydroxyl group of the ligand [74]. It is likely that the presence of intermediate conformational states is generalizable to GPCRs able to bind much larger ligands, i.e. peptides or hormones [75].

10.8.2
Molecular Processes that Propagate the Signal from the Ligand Binding Site to the Intracellular Amino Acids of the Transmembrane Bundle

As detailed above, agonist binding directly affects F6.52 and W6.48 which are part of the conserved aromatic CWxPFF6.52 cluster at the top of TM6 associated with the process of receptor activation [12]. It has recently been observed in the struc-

ture of metarhodopsin I, as determined by electron crystallography [65], that W6.48 undergoes a conformational transition in the process of receptor activation, from pointing towards TM7, in the inactive *gauche+* conformation, to pointing towards TM5, in the active *trans* conformation (Fig. 10.8) as was previously suggested by experimental studies in rhodopsin [76] and computer simulations [73]. Thus, the highly conserved hydrogen bond network linking D2.50 and W6.48 appears to maintain the receptor in the inactive state (Fig. 10.7). For instance, disruption of this network by mutating either N3.35 in the AT1 receptor [77] or N7.45 in the histamine H_1 receptor [66] leads to constitutive activation of the receptor. Rearrangement of W6.48 and the nearby C6.47 decreases the proline-induced bend angle of TM6 [12, 73], moving the cytoplasmic end of TM6 away from TM3 and disrupting the proposed ionic lock (see Fig. 10.6) between TM6 (D/E6.30) and R3.50 of the highly conserved (D/E)RY motif in TM3 [56], aided by the protonation of (D/E3.49) [57, 78, 79]. These large conformational changes in TM3 and TM6 are considered to be an important step in the process of GPCR activation and have also received experimental support from various biophysical and mutational studies [10, 58].

The highly conserved NPxxY motif in TM7 is related to another process involved in the propagation of the signal across the receptor. N7.49 acts as an on/off switch by adopting two different conformations in the inactive and active states, as was shown for the thyrotropin receptor [67, 80]. In the inactive *gauche+* conformation, N7.49 is restrained towards TM6 through the interaction with either the TD6.44 motif, specific for the glycoprotein hormone receptor family [67], or via a water molecule in rhodopsin (and possibly other receptors) (Fig. 10.7 d) [18]. Upon receptor activation N7.49 adopts the *trans* conformation to interact with D2.50 in TM2 (Fig. 10.8) [67]. Due to the presence of the negatively charged D2.50, the O_δ
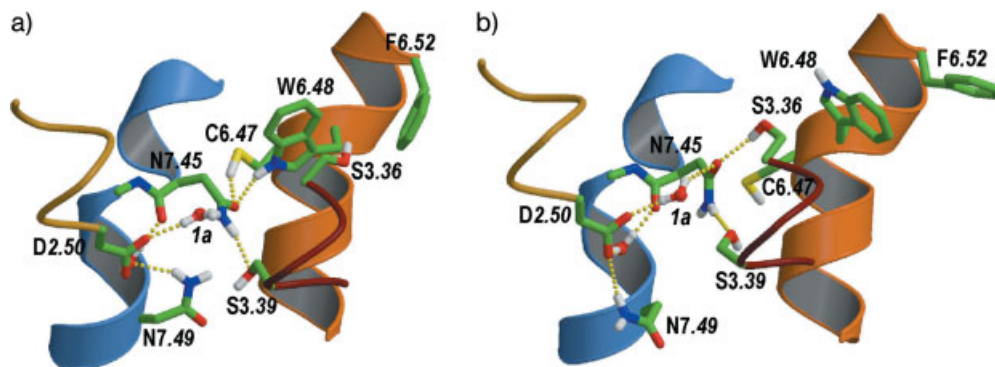
Fig. 10.8 Proposed model of GPCR activation exemplified by the histamine H_1 receptor [66]. The conformation of the C6.47, W6.48, F6.52, N7.45, and N7.49 side chains has been modified from the inactive (a) to the proposed active (b) conformations. It is important to note that receptor activation requires additional breakage and/or formation of interhelical interactions and rigid-body motions of several TM helices. The C_α traces of TMs 2 (golden red), 3 (dark red), 6 (orange), and 7 (blue) are shown. Only polar hydrogens are shown to give a better view.

atom of N7.49 is more polar than a regular carbonyl of neutral Asn. This combination of polar and charged side chains leads to a particular electrostatic landscape which could force the re-orientation of another highly conserved charge, R3.50 in TM3, and drive important conformational reorganization in the TM bundle [67].

Another residue of the NPxxY motif in TM7, Y7.53, interacts with F7.60 in HX8 (Fig. 10.7 e). It has been suggested that this aromatic–aromatic interaction is disrupted during receptor activation, leading to the proper re-aligning of HX8 [81, 82].

The molecular actors detailed above constitute the pieces of a puzzle that is slowly coming together. In particular, these thorough analyses give an indication of the molecular action of the agonist at the receptor. In the light of current knowledge regarding the structure and mechanisms of activation of the GPCRs, we suggest that the following processes may be responsible for the propagation of the signal from the ligand binding site to the intracellular amino acids of the transmembrane bundle: (1) agonist binding directly or indirectly leads to a re-organization of the aromatic cluster in TM6, where F6.52 goes from *trans* to *gauche+*, and W6.48 from *trans* to *gauche+* [65, 73]; (2) N7.49 of the NPxxY motif in TM7 changes from *gauche+* to *trans*, releasing the constraining interaction with TM6 and forming an interaction with D2.50 in TM2 [67, 80]; (3) removal of the ionic lock between R3.50 of the highly conserved (D/E)RY motif in TM3 and a partly conserved D/E6.30 residue in TM6 [56] leads to the outwards rigid body motion of TM6; (4) re-orientation of R3.50 towards the negatively charged D2.50–N7.49 pair; and (5) removal of the aromatic–aromatic Y7.53–F7.60 interaction and re-aligment of HX8. Figure 10.8 shows the computer model where the conformation of these side chains have been modified from the inactive (Fig. 10.8 a) to the proposed active (Fig. 10.8 b) conformations.

10.9
Conclusions

The rhodopsin family of GPCRs is characterized by a number of highly conserved charged and polar residues located within the TM region. Mutagenesis studies indicate that most if not all of these amino acids are involved in maintaining the three-dimensional structure of the receptor and in the processes of receptor activation. As these key residues are common to most of the GPCR members, we hypothesized that these molecular steps form a conserved mechanism.

On the other hand, a detailed analysis of the structure of rhodopsin shows how its transmembrane helices are far from being ideal or even standard Pro-kinked α-helices. We have shown how some of the strong structural distortions are linked to highly conserved residues and, therefore, likely to be conserved throughout the whole Class A GPCRs. Thus, the use of rhodopsin as a template takes into account many of these conserved structural subtleties. Finally, it must be emphasized that homology modeling of GPCRs is far from being a routine technique. But despite its complexity and inherent uncertainties, the inclusion of experimental results and structural specificities for each receptor subfamily can improve the

reliability of the models. These tailor-made models take into account the differences in ligand binding and in how binding is translated into the common steps of receptor activation.

References

1 T. H. Ji, M. Grossmann, and I. Ji. G Protein-coupled receptors. I. Diversity of receptor–ligand interactions. *J Biol Chem* **1998**; *273*: 17299–17302.

2 J. Bockaert, and J. P. Pin. Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* **1999**; *18*: 1723–1729.

3 U. Gether. Uncovering molecular mechanisms involved in activation of G protein-coupled receptors. *Endocr Rev* **2000**; *21*: 90–113.

4 K. Kristiansen. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther* **2004**; *103*: 21–80.

5 J. Drews. Drug discovery: a historical perspective. *Science* **2000**; *287*: 1960–1964.

6 A. L. Hopkins, and C. R. Groom. The druggable genome. *Nat Rev Drug Discov* **2002**; *1*: 727–730.

7 T. Klabunde, and G. Hessler. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **2002**; *3*: 928–944.

8 P. Samama, S. Cotecchia, T. Costa, and R. J. Lefkowitz. A Mutation-induced activated state of the β_2-adrenergic receptor. *J Biol Chem* **1993**; *268*: 4625–4636.

9 R. A. Bond, P. Leff, T. D. Johnson, C. A. Milano, H. A. Rockman, T. R. McMinn, S. Apparsundaram, M. F. Hyek, T. P. Kenakin, L. F. Allen, and R. J. Lefkowitz. Physiological effects of inverse agonists in transgenic mice with myocardial overexpression of the beta(2)-adrenoceptor. *Nature* **1995**; *374*: 272–276.

10 E. C. Meng, and H. R. Bourne. Receptor activation: what does the rhodopsin structure tell us? *Trends Pharmacol Sci* **2001**; *22*: 587–593.

11 U. Gether, and B. K. Kobilka. G Protein-coupled receptors. II. Mechanism of agonist activation. *J Biol Chem* **1998**; *273*: 17979–17982.

12 I. Visiers, J. A. Ballesteros, and H. Weinstein. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol* **2002**; *343*: 329–371.

13 R. J. Lefkowitz, S. Cotecchia, P. Samama, and T. Costa. Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins. *Trends Pharmacol Sci* **1993**; *14*: 303–307.

14 P. Ghanouni, Z. Gryczynski, J. J. Steenhuis, T. W. Lee, D. L. Farrens, J. R. Lakowicz, and B. K. Kobilka. Functionally different agonists induce distinct conformations in the G protein coupling domain of the beta 2 adrenergic receptor. *J Biol Chem* **2001**; *276*: 24433–24436.

15 G. Swaminath, Y. Xiang, T. W. Lee, J. Steenhuis, C. Parnot, and B. K. Kobilka. Sequential binding of agonists to the beta2 adrenoceptor. Kinetic evidence for intermediate conformational states. *J Biol Chem* **2004**; *279*: 686–691.

16 K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. L. Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

17 T. Okada, M. Sugihara, A. N. Bondar, M. Elstner, P. Entel, and V. Buss. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J Mol Biol* **2004**; *342*: 571–583.

18 T. Okada, Y. Fujiyoshi, M. Silow, J. Navarro, E. M. Landau, and Y. Shichida. Functional role of internal water molecules in rhodopsin revealed

by X-ray crystallography. *Proc Natl Acad Sci USA* **2002**; *99*: 5982–5987.

19 J. Li, P. C. Edwards, M. Burghammer, C. Villa, and G. F. Schertler. Structure of bovine rhodopsin in a trigonal crystal form. *J Mol Biol* **2004**; *343*: 1409–1438.

20 T. Okada, and K. Palczewski. Crystal structure of rhodopsin: implications for vision and beyond. *Curr Opin Struct Biol* **2001**; *11*: 420–426.

21 S. Filipek, D. C. Teller, K. Palczewski, and R. Stenkamp. The crystallographic model of rhodopsin and its use in studies of other G protein-coupled receptors. *Annu Rev Biophys Biomol Struct* **2003**; *32*: 375–397.

22 C. Bissantz, P. Bernard, M. Hibert, and D. Rognan. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* **2003**; *50*: 5–25.

23 R. Fredriksson, M. C. Lagerstrom, L. G. Lundin, and H. B. Schioth. The g-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

24 O. Civelli. GPCR deorphanizations: the novel, the known and the unexpected transmitters. *Trends Pharmacol Sci* **2005**; *26*: 15–19.

25 D. K. Vassilatis, J. G. Hohmann, H. Zeng, F. Li, J. E. Ranchalis, M. T. Mortrud, A. Brown, S. S. Rodriguez, J. R. Weller, A. C. Wright, J. E. Bergmann, and G. A. Gaitanaris. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci USA* **2003**; *100*: 4903–4908.

26 N. Kunishima, Y. Shimada, Y. Tsuji, T. Sato, M. Yamamoto, T. Kumasaka, S. Nakanishi, H. Jingami, and K. Morikawa. Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **2000**; *407*: 971–977.

27 A. P. West, Jr., L. L. Llamas, P. M. Snow, S. Benzer, and P. J. Bjorkman. Crystal structure of the ectodomain of Methuselah, a Drosophila G protein-coupled receptor associated with extended life-span. *Proc Natl Acad Sci USA* **2001**; *98*: 3744–3749.

28 C. E. Dann, J. C. Hsieh, A. Rattner, D. Sharma, J. Nathans, and D. J. Leahy. Insights into Wnt binding and signalling from the structures of two Frizzled cysteine-rich domains. *Nature* **2001**; *412*: 86–90.

29 C. R. Grace, M. H. Perrin, M. R. DiGruccio, C. L. Miller, J. E. Rivier, W. W. Vale, and R. Riek. NMR structure and peptide hormone binding site of the first extracellular domain of a type B1 G protein-coupled receptor. *Proc Natl Acad Sci USA* **2004**; *101*: 12836–12841.

30 Q. R. Fan, and W. A. Hendrickson. Structure of human follicle-stimulating hormone in complex with its receptor. *Nature* **2005**; *433*: 269–277.

31 T. Mirzadegan, G. Benko, S. Filipek, and K. Palczewski. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* **2003**; *42*: 2759–2767.

32 J. A. Ballesteros, and H. Weinstein. Integrated methods for the construction of three dimensional models and computational probing of structure–function relations in G-protein coupled receptors. *Methods Neurosci* **1995**; *25*: 366–428.

33 H. R. Bourne, and E. C. Meng. Rhodopsin sees the light. *Science* **2000**; *289*: 733–734.

34 K. T. O'Neil, and W. F. DeGrado. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* **1990**; *250*: 646–651.

35 M. Monne, M. Hermansson, and G. von Heijne. A turn propensity scale for transmembrane helices. *J Mol Biol* **1999**; *288*: 141–145.

36 A. Senes, M. Gerstein, and D. M. Engelman. Statistical analysis of amino acid patterns in transmembrane helices: The GxxG motif occurs frequently and in association with b-branched residues at neighboring positions. *J Mol Biol* **2000**; *296*: 921–936.

37 G. Von Heijne. Proline kinks in transmembrane alpha-helices. *J Mol Biol* **1991**; *218*: 499–503.

38 P. Chakrabarti, and S. Chakrabarti. C–H···O hydrogen bond involving proline residues in alpha-helices. *J Mol Biol* **1998**; *284*: 867–873.

39 F. S. Cordes, J. N. Bright, and M. S. Sansom. Proline-induced distortions of transmembrane helices. *J Mol Biol* **2002**; *323*: 951–960.

40 X. Deupi, M. Olivella, C. Govaerts, J. A. Ballesteros, M. Campillo, and L. Pardo. Ser and Thr residues modulate the conformation of Pro-kinked transmembrane alpha-helices. *Biophys J* **2004**; *86*: 105–115.

41 M. S. P. Sansom, and H. Weinstein. Hinges, swivels and switches: the role of prolines in signalling via transmembrane a-helices. *Trends Pharmacol Sci* **2000**; *21*: 445–451.

42 U. Gether, S. Lin, P. Ghanouni, J. A. Ballesteros, H. Weinstein, and B. K. Kobilka. Agonists induce conformational changes in transmembrane domains III and VI of the β_2 adrenergic receptor. *EMBO J* **1997**; *16*: 6737–6747.

43 I. Visiers, B. B. Braunheim, and H. Weinstein. Prokink: a protocol for numerical evaluation of helix distortions by proline. *Protein Eng* **2000**; *13*: 603–606.

44 M. Bansal, S. Kumar, and R. Velavan. HELANAL: a program to characterize helix geometry in proteins. *J Biomol Struct Dyn* **2000**; *17*: 811–819.

45 J. A. Ballesteros, L. Shi, and J. A. Javitch. Structural mimicry in G Protein-couple receptors: implications of the high-resolution structure of rhodopsin for structure–function analysis of rhodopsin-like receptors. *Mol Pharmacol* **2001**; *60*: 1–19.

46 A. Senes, I. Ubarretxena-Belandia, and D. M. Engelman. The Calpha–H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc Natl Acad Sci USA* **2001**; *98*: 9056–9061.

47 C. Govaerts, C. Blanpain, X. Deupi, S. Ballet, J. A. Ballesteros, S. J. Wodak, G. Vassart, L. Pardo, and M. Parmentier. The TxP motif in the second transmembrane helix of CCR5: a structural determinant in chemokine-induced activation. *J Biol Chem* **2001**; *276*: 13217–13225.

48 M. L. Lopez-Rodriguez, B. Vicente, X. Deupi, S. Barrondo, M. Olivella, M. J. Morcillo, B. Behamu, J. A. Ballesteros, J. Salles, and L. Pardo. Design, synthesis and pharmacological evaluation of 5-hydroxytryptamine(1a) receptor ligands to explore the three-dimensional structure of the receptor. *Mol Pharmacol* **2002**; *62*: 15–21.

49 C. Govaerts, A. Bondue, J. Y. Springael, M. Olivella, X. Deupi, E. Le Poul, S. J. Wodak, M. Parmentier, L. Pardo, and C. Blanpain. Activation of CCR5 by chemokines involves an aromatic cluster between transmembrane helices 2 and 3. *J Biol Chem* **2003**; *278*: 1892–1903.

50 L. Shi, and J. A. Javitch. The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. *Proc Natl Acad Sci USA* **2004**; *101*: 440–445.

51 J. A. Ballesteros, X. Deupi, M. Olivella, E. E. J. Haaksma, and L. Pardo. Serine and threonine residues bend α-helices in the χ_1 = g-conformation. *Biophys J* **2000**; *79*: 2754–2760.

52 L. Shi, and J. A. Javitch. The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu Rev Pharmacol Toxicol* **2002**; *42*: 437–467.

53 J. A. DeMartino, Z. D. Konteatis, S. J. Siciliano, G. Van Riper, D. J. Underwood, P. A. Fischer, and M. S. Springer. Arginine 206 of the C5a receptor is critical for ligand recognition and receptor activation by C-terminal hexapeptide analogs. *J Biol Chem* **1995**; *270*: 15966–15969.

54 J. S. Mills, H. M. Miettinen, D. Cummings, and A. J. Jesaitis. Characterization of the binding site on the formyl peptide receptor using three receptor mutants and analogs of Met-Leu-Phe and Met-Met-Trp-Leu-Leu. *J Biol Chem* **2000**; *275*: 39012–39017.

55 G. Liapakis, J. A. Ballesteros, S. Papachristou, W. C. Chan, X. Chen, and J. A. Javitch. The forgotten serine. A critical role for ser-203(5.42) in ligand

binding and activation of the beta 2-adrenergic receptor. *J Biol Chem* **2000**; *275*: 37779–37788.

56 J. A. Ballesteros, A. D. Jensen, G. Liapakis, S. G. Rasmussen, L. Shi, U. Gether, and J. A. Javitch. Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J Biol Chem* **2001**; *276*: 29171–29177.

57 A. Scheer, F. Fanelli, T. Costa, P. G. De Benedetti, and S. Cotecchia. Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation. *EMBO J* **1996**; *15*: 3566–3578.

58 D. L. Farrens, C. Altenbach, K. Yang, W. L. Hubbell, and H. G. Khorana. Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin. *Science* **1996**; *274*: 768–770.

59 P. Huang, I. Visiers, H. Weinstein, and L. Y. Liu-Chen. The local environment at the cytoplasmic end of TM6 of the mu opioid receptor differs from those of rhodopsin and monoamine receptors: introduction of an ionic lock between the cytoplasmic ends of helices 3 and 6 by a L6.30(275)E mutation inactivates the mu opioid receptor and reduces the constitutive activity of its T6.34(279)K mutant. *Biochemistry* **2002**; *41*: 11972–11980.

60 J. A. Javitch, L. Shi, and G. Liapakis. Use of the substituted cysteine accessibility method to study the structure and function of G protein-coupled receptors. *Methods Enzymol* **2002**; *343*: 137–156.

61 A. A. Boucard, M. Roy, M. E. Beaulieu, P. Lavigne, E. Escher, G. Guillemette, and R. Leduc. Constitutive activation of the angiotensin II type 1 receptor alters the spatial proximity of transmembrane 7 to the ligand-binding pocket. *J Biol Chem* **2003**; *278*: 36628–36636.

62 W. Xu, M. Campillo, L. Pardo, J. K. de Riel, and L.-Y. Liu-Chen. The seventh transmembrane domains of the delta and kappa opioid receptors have

different accessibility patterns and interhelical interactions. *Biochemistry* **2005** (in press).

63 D. Fu, J. A. Ballesteros, H. Weinstein, J. Chen, and J. A. Javitch. Residues in the seventh membrane-spanning segment of the dopamine D_2 receptor accessible in the binding-site crevice. *Biochemistry* **1996**; *35*: 11278–11285.

64 E. S. Dawson, and J. N. Wells. Determination of amino acid residues that are accessible from the ligand binding crevice in the seventh transmembrane-spanning region of the human A(1) adenosine receptor. *Mol Pharmacol* **2001**; *59*: 1187–1195.

65 J. J. Ruprecht, T. Mielke, R. Vogel, C. Villa, and G. F. Schertler. Electron crystallography reveals the structure of metarhodopsin I. *EMBO J* **2004**; *23*: 3609–3620.

66 A. Jongejan, M. Bruysters, J. A. Ballesteros, E. Haaksma, R. A. Bakker, L. Pardo, and R. Leurs. Linking ligand binding to histamine H_1 receptor activation. *Nat Chem Biol* **2005**; *1*: 98–103.

67 E. Urizar, S. Claeysen, X. Deupi, C. Govaerts, S. Costagliola, G. Vassart, and L. Pardo. An activation switch in the rhodopsin family of G protein-coupled receptors: the thyrotropin receptor. *J Biol Chem* **2005**; *280*: 17135–17141.

68 K. D. Ridge, N. G. Abdulaev, M. Sousa, and K. Palczewski. Phototransduction: crystal clear. *Trends Biochem Sci* **2003**; *28*: 479–487.

69 P. Ghanouni, J. J. Steenhuis, D. L. Farrens, and B. K. Kobilka. Agonist-induced conformational changes in the G-protein-coupling domain of the beta 2 adrenergic receptor. *Proc Natl Acad Sci USA* **2001**; *98*: 5997–6002.

70 B. Kobilka. Agonist binding: a multi-step process. *Mol Pharmacol* **2004**; *65*: 1060–1062.

71 C. D. Strader, I. S. Sigal, and R. A. F. Dixon. Structural basis of β-adrenergic receptor function. *FASEB J* **1989**; *3*: 1825–1832.

72 C. D. Strader, M. R. Candelore, W. S. Hill, I. S. Sigal, and R. A. F. Dixon. Identification of two serine residues involved in agonist activation of the â-adrenergic receptor. *J Biol Chem* **1989**; *264*: 13572–13578.

73 L. Shi, G. Liapakis, R. Xu, F. Guarnieri, J. A. Ballesteros, and J. A. Javitch. Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J Biol Chem* **2002**; *277*: 40989–40996.

74 K. Wieland, H. M. Zuurmond, C. Krasel, A. P. Ijzerman, and M. J. Lohse. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc Natl Acad Sci USA* **1996**; *93*: 9276–9281.

75 M. Azzi, P. G. Charest, S. Angers, G. Rousseau, T. Kohout, M. Bouvier, and G. Pineyro. Beta-arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proc Natl Acad Sci USA* **2003**; *100*: 11406–11411.

76 S. W. Lin, and T. P. Sakmar. Specific tryptophan UV-absorbance changes are probes of the transition of rhodopsin to its active state. *Biochemistry* **1996**; *35*: 11149–11159.

77 M. Auger-Messier, M. Clement, P. M. Lanctot, P. C. Leclerc, R. Leduc, E. Escher, and G. Guillemette. The constitutively active N111G-AT1 receptor for angiotensin II maintains a high affinity conformation despite being uncoupled from its cognate G protein Gq/11alpha. *Endocrinology* **2003**; *144*: 5277–5284.

78 L. Oliveira, A. C. M. Paiva, C. Sander, and G. Vriend. A common step for signal transduction in G-protein coupled receptors. *Trends Pharmacol Sci* **1994**; *15*: 170–172.

79 P. Ghanouni, H. Schambye, R. Seifert, T. W. Lee, S. G. Rasmussen, U. Gether, and B. K. Kobilka. The effect of pH on beta(2) adrenoceptor function. Evidence for protonation-dependent activation. *J Biol Chem* **2000**; *275*: 3121–3127.

80 C. Govaerts, A. Lefort, S. Costagliola, S. Wodak, J. A. Ballesteros, L. Pardo, and G. Vassart. A conserved Asn in TM7 is an on/off switch in the activation of the TSH receptor. *J Biol Chem* **2001**; *276*: 22991–22999.

81 C. Prioleau, I. Visiers, B. J. Ebersole, H. Weinstein, and S. C. Sealfon. Conserved helix 7 tyrosine acts as a multistate conformational switch in the 5HT2C receptor. identification of a novel "locked-on" phenotype and double revertant mutations. *J Biol Chem* **2002**; *277*: 36577–36584.

82 O. Fritze, S. Filipek, V. Kuksa, K. Palczewski, K. P. Hofmann, and O. P. Ernst. Role of the conserved NPxxY(x)5,6F motif in the rhodopsin ground state and during activation. *Proc Natl Acad Sci USA* **2003**; *100*: 2290–2295.

83 J. Kraulis. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J Appl Cryst* **1991**; *24*: 946–950.

84 E. A. Merritt, and D. J. Bacon. Raster3D: photorealistic molecular graphics. *Methods Enzymol* **1997**; *277*: 505–524.

11
7TM Models in Structure-based Drug Design

Frank E. Blaney, Anna-Maria Capelli, and Giovanna Tedesco

11.1
Introduction

The G protein-coupled receptor or as it is commonly known, 7TM receptor family has become one of the most widely studied protein classes in existence today. This is due to its extensive involvement in modern drug therapy; indeed it has been estimated that more than 50% of known drugs act on these receptors. The family can be divided into five or more subclasses of which three are found in humans and have been the target of therapeutic studies. By far the largest is the rhodopsin-like ("Family A") group. A vast variety of stimuli can specifically activate these proteins including light, taste and odorant molecules, monoaminergic neurotransmitters, peptide hormones, glycoproteins, chemokines, nucleotides, lipids and prostaglandins. Although they are fewer in number, "Family B" (secretin-like) receptors are the targets of a number of important peptide hormones. They are characterized by a large N-terminal domain containing three conserved disulfide bonds. In "Family C" the competitive agonist ligands bind exclusively in the very large (900–1200 residues) N-terminal domain which is similar in structure to the periplasmic binding proteins. This bilobal structure then switches from an open to closed state presumably causing some allosteric conformational change in the TM domain. Relatively few members of this family are known; among them being the metabotropic glutamate, the GABA-B and the calcium sensing receptors.

Medicinal chemistry programs targeted towards 7TM receptors typically involve the syntheses of both (partial) agonist and/or antagonist compounds. The former however are generally considered more difficult to achieve as these syntheses involve a different conformational state of the receptor for which little experimental data are known. More recently allosteric modulators have become important as drug targets for GPCRs. The importance of these receptors has led to widespread efforts in the modeling of them. These models have undoubtedly improved and in the pharmaceutical industry they are widely used for *in silico* searches of virtual databases or internal compound collections composed of ligands previously developed for other diseases as well as combinatorial libraries or purchased sets. They
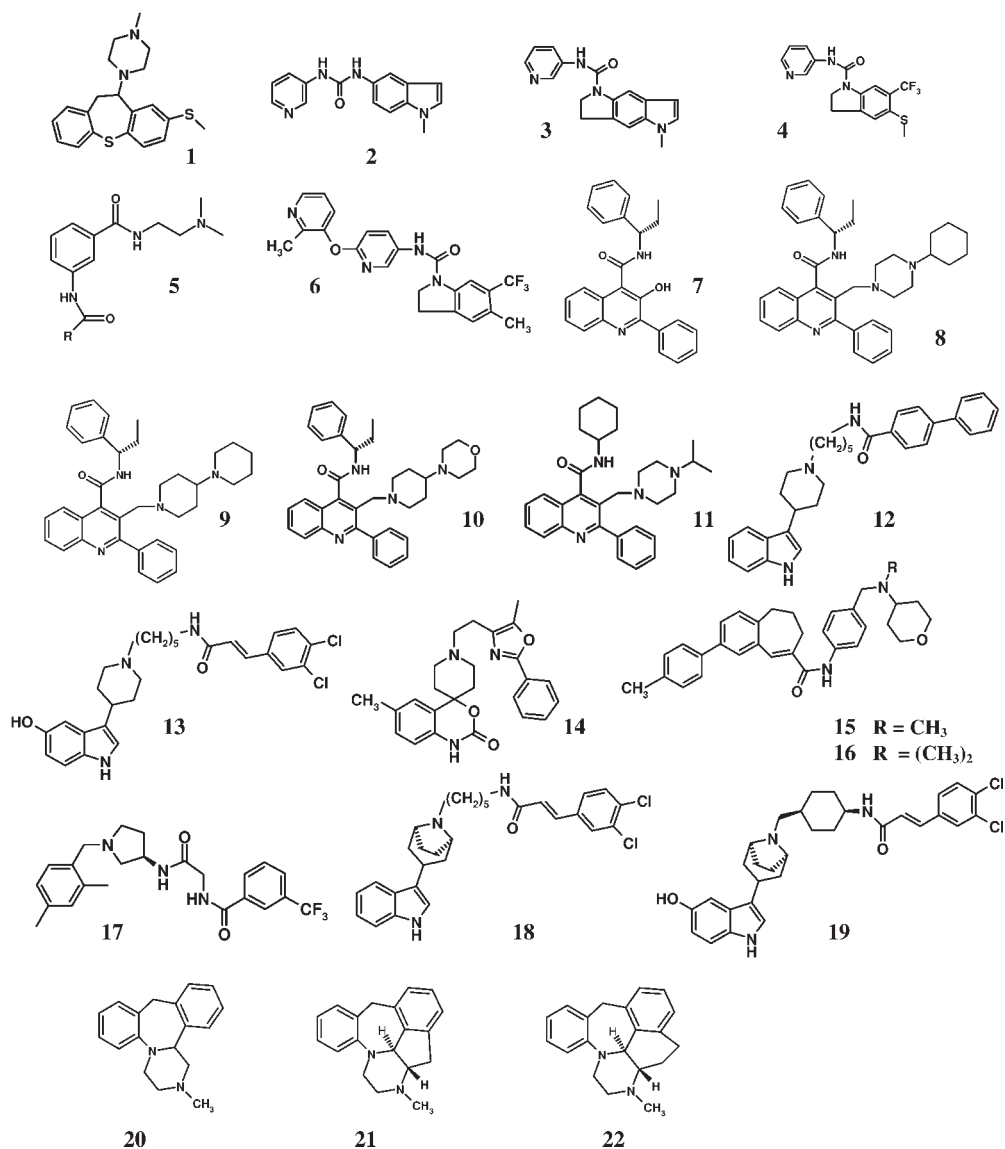
Fig. 11.1 Compounds referred to in the text.
1 methiophepin, a classic tricyclic drug;
2 SB-200646; **3** SB-206553; **4** SB-221284;
5 a series of compounds with varying substituted dihydroindole ring R groups; **6** SB-243213; **2–6** are selective 5-HT$_{2C}$ antagonists.
7 SB-223412 or Talnetant, a selective NK3 antagonist; **8** SB-327076, potent NK2/NK3 mixed antagonist; **9** SB-371059, mixed NK2/NK3 antagonist devoid of μ-opioid activity;

10 SB-400238, the most potent NK2/NK3 antagonist; **11** SB-415707, selective NK2 antagonist; **12** a CCR2 lead from a 5-HT series; **13** SB-282241; **14** RS-504393; **15** Takeda free base; **16** TAK-779; **17** Teijin lead compound; **18** tropane analog of SB-282241; **19** SB-380732; **12–19** are potent CCR2 antagonists. **20** mianserin; **21** BRL34849; **22** BRL34969; **20–22** are 5-HT$_6$ antagonists.

are also frequently used to explain experimental results or in the design and rationalization of SDM experiments. Despite their evident value relatively little has been reported on their use in structure-based drug design. Many people will argue that only real crystal structures are of use here and, as these are only models, they are of limited interest. This is a disappointing attitude because it could be argued that with a crystal structure, but without a bound ligand, it is equally difficult to carry out structure-based design. A good starting protein structure is only the first step in the design process. It is then necessary to dock known ligands to identify potential binding sites. These hypothetical sites should then be tested by experimental ligand SAR and site-directed mutagenesis work where possible. Only then can the design phase be implemented. It is the ligand docking phase which is crucial to the whole procedure. In the following pages a description of how the protein models have evolved and improved will be given. Examples will be shown where they have been successfully used in the design process.

11.2
Early Models of 7TM Receptors

Towards the very late 1980s through to 1991 several groups had started to construct molecular models of 7TM receptors [2, 3]. Essentially two approaches were used. The more common one was to assume that there was some perceived homology with the non-G protein-coupled 7TM membrane protein, bacteriorhodopsin [4], for which a three-dimensional structure was known. While such homology is debatable, it is well known that proteins with no sequence identity can often adopt the same fold. It was reasonable to assume therefore that the GPCR receptor families which were also assumed to contain 7 TM helices would pack in a similar manner to bacteriorhodopsin (BR). Many groups started to use BR as a template but perhaps the best known of these was that of Hibert et al. [5]. In their pioneering paper for example, they showed that the catecholamine agonist ligands such as dopamine and noradrenaline, were able to bind their primary amino headgroup to a conserved aspartate on TM3, while at the same time forming hydrogen bonds between the catechol hydroxyls and serines on TM5. This binding mode was in agreement with early SDM work in the β2 adrenergic receptor [6]. An aromatic cage formed by residues on TM5 and TM6 was also postulated and this is still a feature of most of the current models today.

The second approach used in early models was to reason that because no homology existed between 7TMs and bacteriorhodopsin, it would be better to build the models using some *de novo* helix packing methodology. The helical regions themselves were usually predicted by hydropathy, using one of the well known scales such as Kyte–Dolittle [7] or Goldman–Engelman–Steitz [8] (Fig. 11.2), while the outer (and hence inner) faces of the helices could be predicted with the use of hydrophobic moments. Packing was often carried out manually, usually followed by molecular mechanics/dynamics simulations for refinement. The main problem was prediction of the inter-helical tilts, something
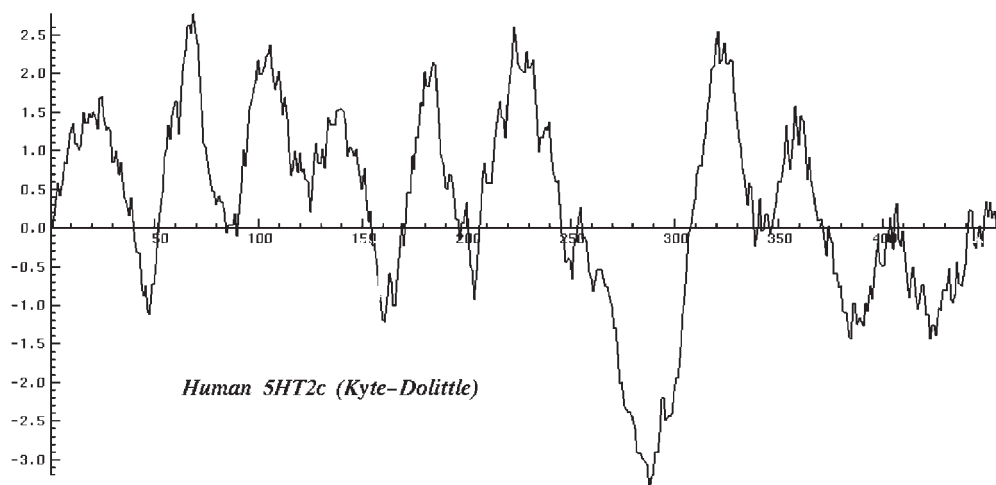
Fig. 11.2 A Kyte–Dolittle hydropathy plot of the 5-HT$_{2C}$ receptor. Note that the first peak is a signal peptide sequence and not a true TM domain.

not readily amenable to standard simulation methods. Vriend compared the inter-helical tilts of membrane-bound helices with those in globular proteins and concluded that there was no difference between the two. They were therefore happy to use BR as a model for their helix packing [9]. In our group we attempted to pack the bundle by maximization of the hydrophobic energy. The reasoning was that the helices would pack in such a way as to maximize their interactions with the lipid bilayer. The Eisenberg method of calculating the hydrophobic energy was used [10]; in this, the energy is the sum over all atoms of the product of each atom's atomic solvation parameter and its accessible surface area. It was found however that the helices tended to drift apart and remain essentially parallel, and the method was abandoned. These early models represented a useful phase in that they established the acceptance of such models within the scientific community, both in academia and in particular in the pharmaceutical area. They were often used successfully to explain known SAR or suggest potential binding modes, but were generally too crude to be of use in real structure-based drug design.

11.3
Third Generation 7TM Models

A "major breakthrough" occurred in 1993 when Schertler published his 9 Å cryo-electron diffraction map of rhodopsin, which was of course a true G protein-coupled receptor [11]. Even at this low resolution it was evident that the packing of the helices in rhodopsin was quite different from that of bacteriorhodopsin, suggesting that the latter was not a viable structure on which to base modeling. Hibert briefly argued that the observed differences in the two cryo-EM maps could

be explained by tilting the whole bundle by approximately 15° in the plane of the bilayer [12]. This was not generally accepted however and work started everywhere to rebuild models based on the new information. The TM domains were still identified by hydropathic analysis but use could also be made of the fact that each domain contained highly conserved key motifs such as the D(E)RY sequence at the bottom of TM3, a trp in TM4, pros in TM5 and TM6 and an NP motif in TM7. The "outward-facing" residues of the bundle are usually hydrophobic and could be identified by using hydrophobic moment calculations. A similar technique developed by Donnolly [13] made use of the fact that the most conserved (often polar or aromatic) residues are those which will be involved in function or binding and will therefore be inward facing. He made use of "conservation moments" to define the inward faces of each helix. Predicting the helix–loop interface points or the inter-helix tilt angles are more difficult problems and remain largely unsolved. Many techniques such as change in moment directions [14], calculation of multiple hydrophobic arcs [15], use of distance constraints coupled with molecular mechanics minimization or dynamics [16, 17] or use of genetic algorithms [18] have been applied to these problems but a full description of them is beyond the scope of this chapter. In our group, a combination of all these tools was used in the construction of the "third generation models". It was really at this time that these models were used in the structure-based design process and continuous refinement of them was possible in light of emerging ligand SAR and SDM studies.

11.3.1
Docking Ligands into Receptor Models

Ligand docking into receptor or enzyme structures has been a major topic of research for many years. Due to its importance in the drug design field a considerable number of programs have been developed for the automated docking of ligands into these structures. Such programs have the advantage of being much faster than manual docking. For the reasons of speed however, they often suffer from the problem of assuming a fixed protein structure, without allowing for the influence of side-chain flexibility or ligand-induced fit. Side-chain flexibility is permitted with a few programs [19] but none allow any major movement of the backbone, a factor frequently observed from crystal structures. They also use some simplified scoring functions to rank dockings rather than a computationally more expensive full energy minimization with potential inclusion of entropy terms. We have used many of these programs and while they are fast and can therefore handle hundreds or even thousands of compounds in a reasonable time, they usually work best for systems with large accessible binding sites, a situation not commonly found in 7TM receptor models. Manual docking on the other hand allows complete freedom of placement of each ligand in the light of knowledge of experimental results as well as the use of optimized ligand geometries, proper *ab initio* electrostatic descriptors and full controlled flexibility of both the ligand and the protein. This has therefore been the method of choice in all the work described here.

11.3.2
Designing 5-HT$_{2C}$ Selective Antagonists

For a long time 5-HT$_{2C}$ has been a target of interest and in fact, this was the first receptor model which was constructed by our group back in 1991. Several lines of evidence suggested that selective 5-HT$_{2C}$ antagonists may be useful in the treatment of anxiety and depression. For example, the moderately selective 5-HT$_{2C}$ agonist mCPP (*meta*-chlorophenylpiperazine), a metabolite of the antidepressant tradozone, causes behavioral indications of anxiety in both animal models and humans. More recently, published data has suggested that 5-HT$_{2C}$ antagonists may offer interesting therapeutic opportunities for the treatment of a variety of CNS disorders such as depression, schizophrenia, migraine and Parkinson's disease.

The main problem in this work was to be able to design compounds which would be selective for 5-HT$_{2C}$ over 5-HT$_{2A}$ receptors. The latter are widely distributed in the periphery as well as the CNS and antagonism of them would give rise to undesirable side effects. Docking studies were carried out on a number of classical non-selective antagonists such as the tricyclics and butyrophenones. It was found that they all occupied a pocket bounded by phe_5:47, phe_6:51 and phe_6:52 and trp_6:48, with the protonated nitrogen forming a salt bridge with the highly conserved asp_3:32 (see Fig. 11.3a). The residues forming this binding pocket are conserved across all 5-HT 7TM receptors and designing selectivity was going to be extremely difficult.

An interesting weakly selective lead compound, SB-200646 (**2**), was however discovered from cross screening and this provided a starting point for further rational design. Although quite weak SB-200646 importantly contained no basic center which suggested immediately that it was not binding to the normal pocket. A study of the SAR showed that the carbonyl of the urea was essential for activity but the NHs were not. In fact cyclization of the urea to give SB-206553 (**3**) led to an increase in binding of one order of magnitude. Numerous analogs of SB-206553 were synthesized and to aid in the understanding of how these might selectively bind to the 5-HT$_{2C}$ receptor, an "active analog" approach was used [20]. This pharmacophoric type method relied on an overlay of the compounds which was achieved by alignment of the urea carbonyls and the aromatic rings. It was hoped that by analysis of the 3D overlay and the associated biological activity, a portion of the molecules would be identified which would explain the observed selectivity for 5-HT$_{2C}$ over 5-HT$_{2A}$. Figure 11.3b shows that indeed such an area was found. This was a small excluded volume on either side of the tricyclic ring which was "allowed" in the 5-HT$_{2C}$ receptor but "disallowed" in 5-HT$_{2A}$. With this in mind docking studies were carried out in the receptor models.

The best solutions found, suggested that the urea carbonyl formed two hydrogen bonds with ser_3:36 and ser_3:39, which are one and two helical turns respectively below the conserved binding aspartate on TM3. The aromatic moieties could sit on either side above and below the urea but energetically the arrangement with the tricyclic ring below the serines was the more favorable. This led to a conformation where the 5-HT$_{2C}$ allowed/5-HT$_{2A}$ disallowed volume was bound
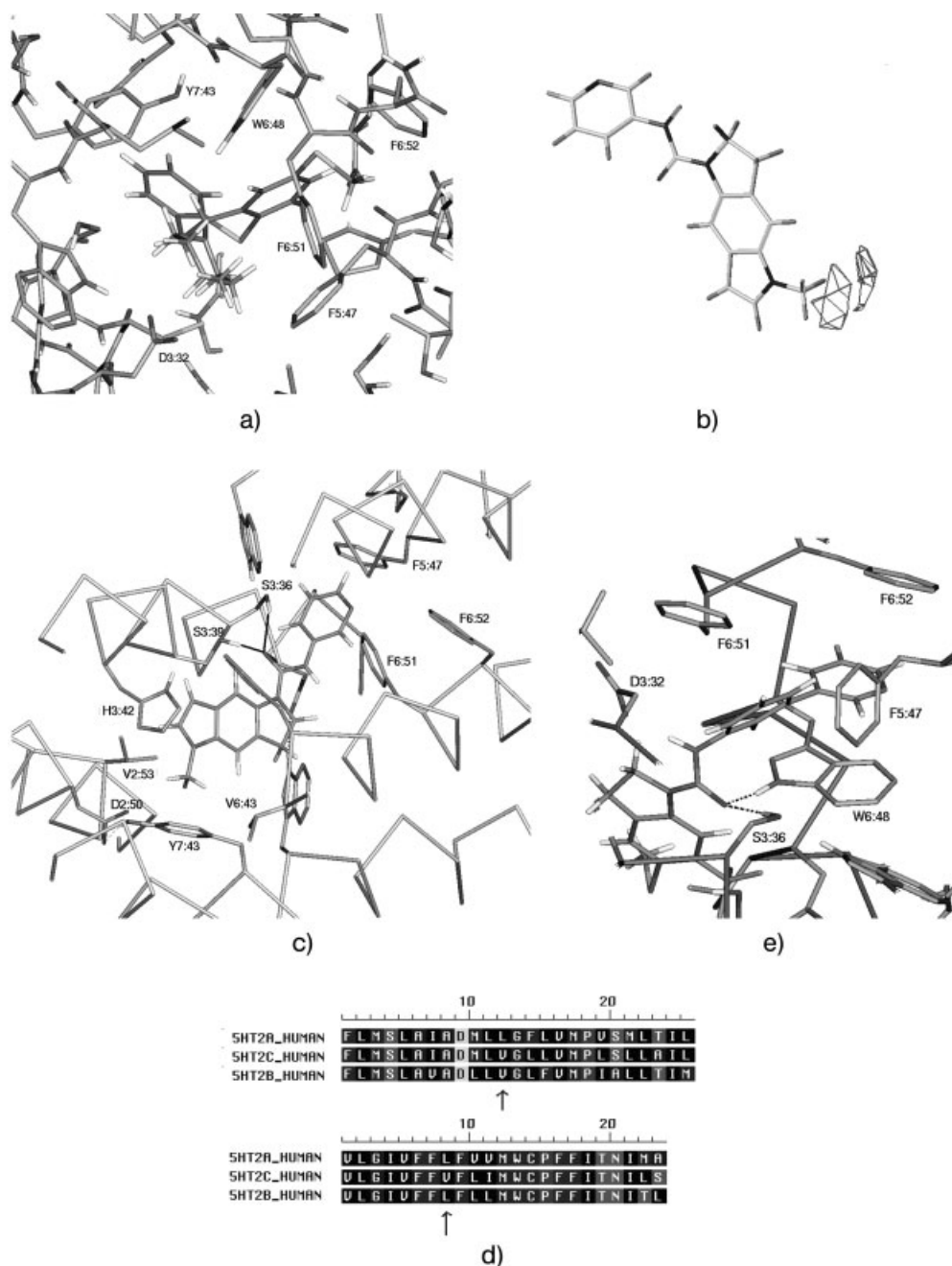
Fig. 11.3 Early 5-HT$_{2C}$ design. (a) Methiophepin docked into the "third generation" 5-HT$_{2C}$ receptor; (b) an active analog model based on SB-206553 analogs; (c) SB-206553 into the 5-HT$_{2C}$ receptor (d) and alignment of 5-HT$_2$ receptors showing TM2 (above) and TM6 (below); (e) the TM5–TM6 aromatic cage with a biaryl compound docked in.

by two valines, val_2:53 and val_6:43 (Fig. 11.3c). It was very exciting to find that the two corresponding residues in 5-HT$_{2A}$ were leucines (Fig. 11.3d) and furthermore that the excluded volume from the active analog model was exactly what would have been expected for a valine–leucine difference.

The first rationalization was that rather than using the tricyclic ring system, it would be better (sterically and synthetically) if the excluded volume was occupied with a suitable group extending from the corresponding bicyclic ring. Either a methoxy or even better a thiomethyl group was deemed ideal and this would be even better if it was forced to sit out-of-plane of the ring (at least on one side) by the introduction of an ortho substituent. This led to the synthesis of SB-221284 (**4**) which was the first ligand to show over 100-fold selectivity for 5-HT$_{2C}$ [21]. The compound was still fairly weak as an antagonist so further elaboration was necessary. Docking of SB-221284 showed that the pyridine ring was placed at the entrance to the aromatic pocket formed by the residues of TM5 and TM6. Extension of the pyridine to a biaryl moiety should fill this cavity so a number of biaryl analogs were synthesized which did indeed increase the potency of the series without loss of selectivity (Fig. 11.4e) [22].

While potency and selectivity had been achieved, the compounds were very insoluble. It was obvious that introduction of a charged amino group would not only improve this but should also lead to an increase in brain penetration. The docking model was studied and it suggested that the inclusion of a four-atom chain branching off the first ring of the biaryl and containing a charged amino group, would be the ideal size to form a salt bridge with the conserved TM3 aspartate. A series of these compounds was synthesized (**5**) and proved to be the first sub-nanomolar selective antagonists of 5-HT$_{2C}$.

11.3.3
Even Wrong Models Can be Useful: The Importance of SDM Studies

The above case study of selective 5-HT$_{2C}$ antagonist design based on receptor modeling is perhaps surprising given the lack of real knowledge about the receptor 3D structure. Given this model however, it was the only docking which explained so elegantly, the results of the active analog approach and which also gave rise to the design of such selective and potent ligands. We were therefore anxious to provide further experimental proof for the validity of the receptor model. One commonly used method of testing a binding hypothesis is to carry out site-directed mutagenesis on residues which are believed to be involved in the receptor pocket. Thus in this case mutation of the two TM3 serines, 3:26 and 3:29, to non hydrogen bonding residues would be expected to decrease binding of the compounds. Both residues were mutated to cysteines and the binding of SB-206553 was studied. It was found that both individual S → C mutants resulted in a slight *increase* in binding of the compound. This of course implied that the receptor model was wrong and that the success of the structure-based design may well have been for the wrong reasons. Unfortunately at the time no other binding hypothesis could be generated.
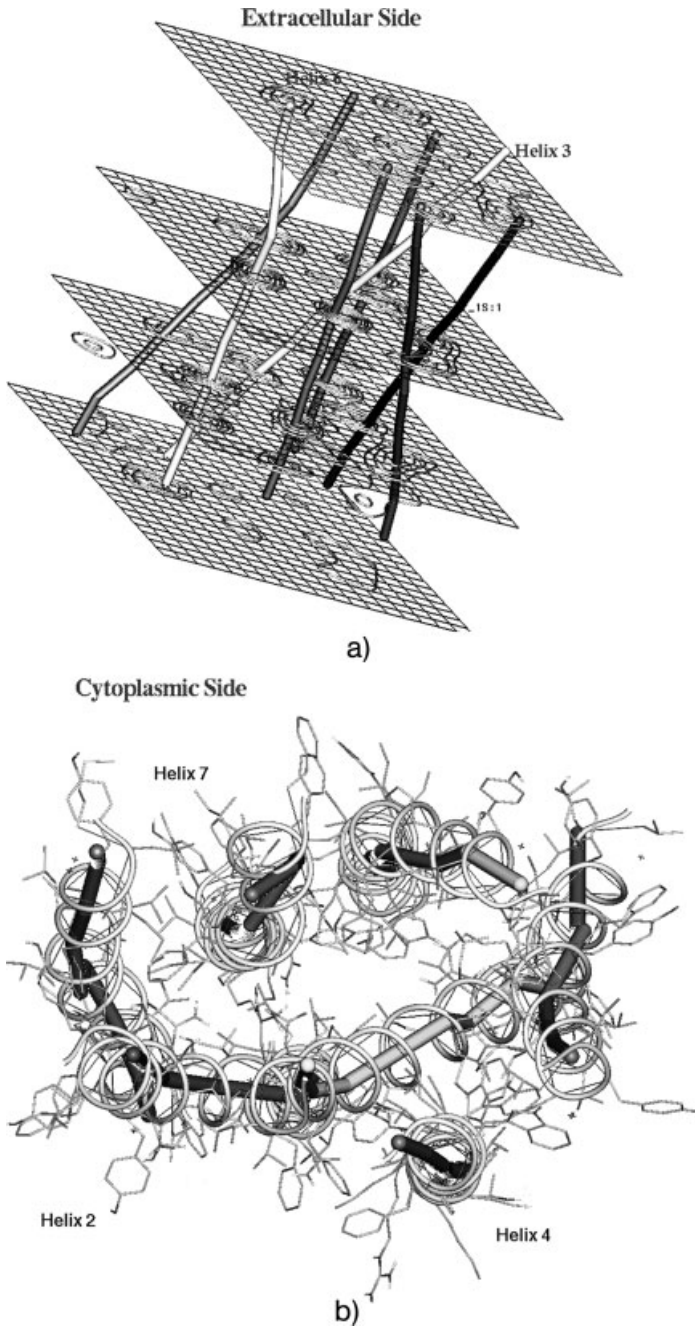
Extracellular Side

a)

Cytoplasmic Side

b)

Fig. 11.4 Construction of a "fourth generation" model. (a) Density slices
through the bilayer with the helical axes constructed through them;
(b) a final model of bovine rhodopsin with the helices folded around the axes.

11.4
Fourth Generation Models

The most exciting publication of the 1990s from the 7TM receptor modeling view-point appeared in 1997 [23], when Schertler described the 6-Å resolution cryoelectron diffraction data for frog rhodopsin at 4-Å slices throughout the bilayer. For the first time therefore it was possible to deduce the interhelical tilts by following the axis of each helix through the density slices. A computer program was written which allowed the conversion of the density pictures in the original papers into real three-dimensional objects. The density maps were then used to generate the helical axes (Fig. 11.4a). The individual helices were allowed to equilibrate in a water–ethane bilayer and were then refolded around these axes. The positioning of TM6 and TM5 was facilitated by the obvious kinks in the helical axes arising from the well-conserved prolines. For the other helices use was made of the HELANAL algorithm [15] together with a wealth of other biophysical and biochemical data such as SDM, SCAM, NMR of retinal binding, engineered metal binding sites, etc. Using all this information a detailed 3D model of the bovine rhodopsin receptor was constructed (Fig. 11.4b) which could be used as a template for other 7TM receptors of interest.

11.4.1
Revisiting the 5-HT$_{2C}$ Antagonist Binding Site

One obvious difference between earlier 7TM models and those derived from the density slice information was the steep angle adopted by TM3 in the bundle. While it "starts" between TM2 and TM4 on the extracellular side as expected, it emerges on the cytoplasmic side of the bundle between TM4 and TM5. The hydrophobic surface potential method described earlier had suggested that the intracellular side would be more buried than the extracellular end, although not to the extent actually found! A consequence of this was that the two serines on TM3 which were originally believed to be involved in binding were no longer accessible to the ligand (Fig. 11.5a). The re-docking of SB-221284 into the new model suggested a somewhat different binding mode. The ligand in both receptors formed a hydrogen bond with a serine in TM4 (4:57; Fig. 11.5a). The pocket was similar to that in the previous generation model with the "top" half of the ligand in proximity to the same aromatic pocket formed by TM5 and TM6. However the binding of the urea carbonyl is to TM4 rather than TM3 and the two valine residues are no longer accessible. There is an alanine on TM5 (ala_5:46) in the same region as the carbonyl binding site. This is present in human 5-HT$_{2C}$ but is a serine in 5-HT$_{2A}$. It is also a serine in the rat 5-HT$_{2C}$ and it has been reported previously that mutation of the human A5:46 S gives rise to a switch in human to rat type pharmacology [24]. It was found in our hands that this mutation also gave rise to a 50-fold loss of binding of SB-206553. Careful docking and full minimization of the receptor–ligand complexes showed that although the positions in the two receptors were similar, in 5-HT$_{2A}$, a strong hydrogen bond between ser_4:57 and
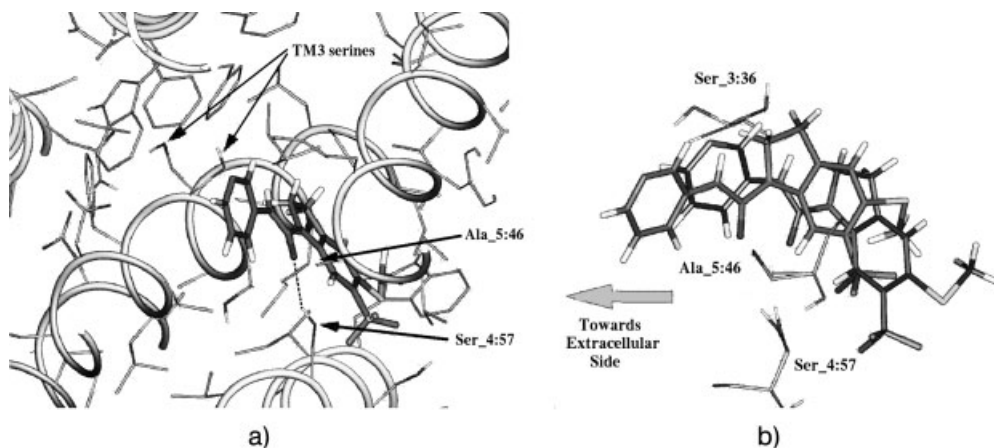
Fig. 11.5 (a) New docking of SB-221284 in the fourth generation 5-HT$_{2C}$ model showing the carbonyl interacting with ser_4:57. The TM3 serines originally implicated are no longer accessible; (b) overlay of SB-221284 as docked into the 5HT$_{2A/2C}$ receptors. In the former it is pushed about 1.5 Å towards the extracellular side.

5:46 led to a shift of about 1.7 Å towards the extracellular side of the TM bundle (Fig. 11.5 b). This shift corresponded to the excluded volume found in the active analog model and thus gave an alternative explanation for it. It also highlighted at the time that it cannot be assumed that the same ligand will bind in an identical manner to two closely related receptors or that a common binding mode of a series of analogs can necessarily be exactly overlaid, a practice which is common to many pharmacophoric designs.

The new docking placed the pyridine ring of SB-221284 in a slightly different position with respect to the aromatic binding pocket and suggested that a one-atom linker between the two rings of the biaryl series would give rise to a better docking geometry (Fig. 11.6 a). The synthesis of SB-243213 (**6**) showed that this was indeed correct as the compound displayed even greater potency/selectivity. Thus once again the use of careful docking followed by the search for additional new potential binding features led to a successful example of structure-based design [25].

11.5
The Inclusion of Extracellular Loops in 7TM Models

In the field of 7TM receptors there is often some component of ligand binding to the extracellular loops. This is particularly true of peptide receptors where SDM and chimeric studies have shown the importance of these regions. Using homology modeling, loops of up to about nine residues have traditionally been generated by searching loop libraries derived from the Brookhaven protein databank.
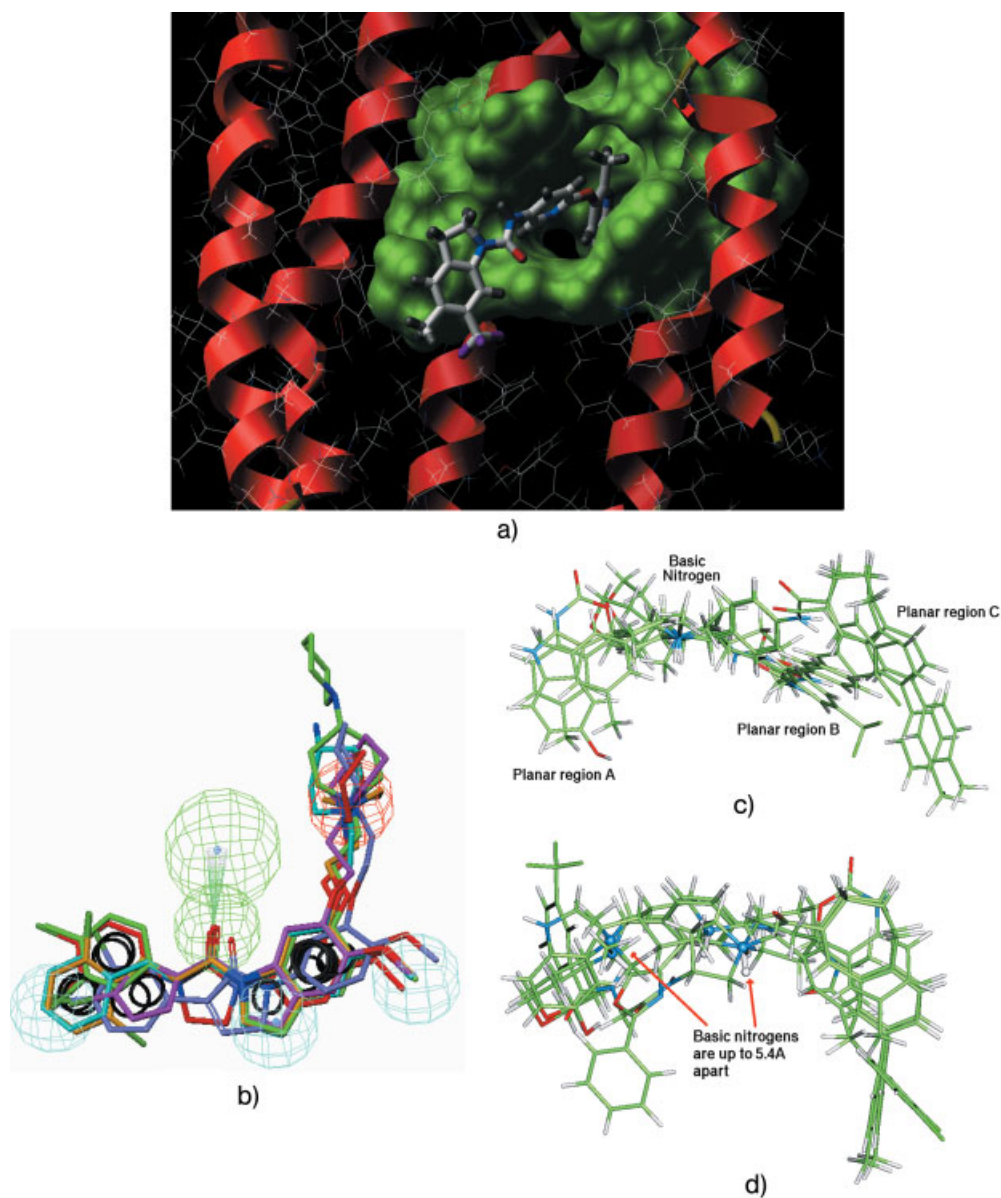
Fig. 11.6 (a) The aromatic cage between TM5 and TM6 of the 5-HT$_{2C}$ model showing SB-243213 docked into it; (b) a classic Catalyst pharmacophore model based on 5-HT$_{2C}$ antagonists. Colour coding of pharmacophoric features: green H-bond acceptor; red, positive ionizable; cyan, hydrophobic. (c) A pharmacophore for CCR2 antagonists based on manual or Catalyst overlays; (d) overlay of the same CCR2 antagonists as they are docked into the receptor model.

However, the extracellular loops of 7TM receptors are often much longer than this and are further complicated by the presence of a highly conserved disulfide bond between TM3 and the second extracellular loop. A semi-automated procedure for the generation of these extracellular loops has been developed in-house and has been widely used in these and subsequent models. In this method initial conformational sampling is performed using the distance geometry program, DGEOM95 [26]. Distance constraints are automatically derived from the five helical residues at each N- and C-terminus of the loop. Further constraints come from the disulfide bond information in ECL2, the proximity of previously generated adjacent loops and consensus secondary structure prediction information. For example if a typical stretch of loop was predicted to exist as a beta-strand using several different prediction methods such as Chou-Fasman, Garnier-Osguthorpe-Robson and Predator, then consensus constraints for the strand backbone distances could automatically be set up; the procedure is shown in Fig. 11.7.
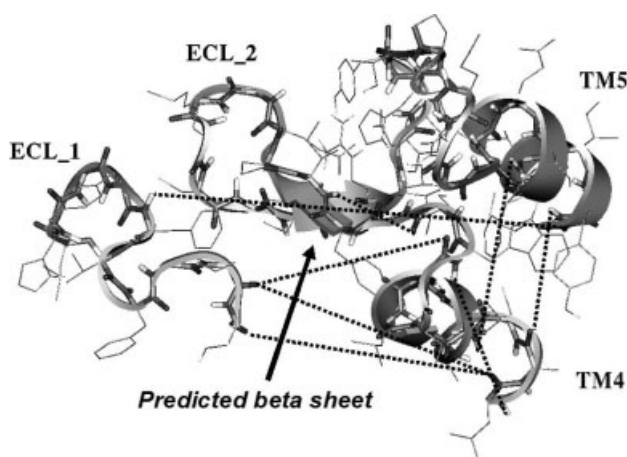
Fig. 11.7 Construction of an extracellular loop region using distance constraints.

Typically 500–1000 random conformations are generated. Each loop structure is minimized with the molecular mechanics program, CHARMm [27] and the conformations are analyzed by goodness of backbone phi–psi angles using the criteria described by Wilmot et al. [28]. They are clustered by backbone dihedral "families" and further ranked by energy. The best conformations were chosen for subsequent molecular dynamics analysis using the locally enhanced sampling (LES) method of Elber [29], where multiple copies are simulated simultaneously until some structural convergence is reached. The best resulting conformations were further minimized before inclusion in the receptor model. While the next generation of 7TM models described below utilized the crystal structure of bovine rhodopsin, there is still an equal uncertainty about the loop structure and the method is still widely used within our group for *de novo* loop modeling. In several in-

stances we have made predictions of loop residues to be involved in binding and these have been subsequently supported by SDM studies [30].

11.6
Switching Selectivity in Neurokinin Antagonists

As was seen with the design of selective 5-HT$_{2C}$ antagonists, one of the greatest challenges that the modeling chemist often encounters is the recognition of subtle features between receptors which are very close in sequence. For example different tissue distributions can exist in subtypes of receptor families and one may wish to target only one of these. Another problem which often arises is that an unwanted side effect due to activity in another unrelated receptor may be found. Designing out one unwanted activity whilst retaining another desired activity, can be a real challenge to the medicinal chemist. One of our most successful examples of fourth generation 7TM model-based design was the switching of selectivity in a neurokinin NK3 compound to a totally selective NK2 analog and to an equipotent mixed NK2/NK3 series.

The neurokinin (tachykinin) receptor family consists of three members, namely NK1, NK2 and NK3, which are activated by the natural peptides substance P, NKA and NKB respectively. The former is an undecapeptide while the latter two are nonapeptides. They share a common Phe-X-Leu_Met-CONH2 motif at the C-terminus. The three receptors are widely distributed in the central (NK1 and NK3) and peripheral nervous systems. Biological actions include smooth muscle contraction and relaxation, vasodilation, secretion (GI and airways), activation of the immune system, pain transmission and neurogenic inflammation. They have been implicated in a broad range of CNS disorders including Parkinson's and Huntington's disease and anxiety and/or depression. Therapeutically, antagonists have been tested for emesis, depression/anxiety and pain/migraine in the CNS and for asthma and GI motility disorders in the periphery. There is still some debate however as to the best profile of selectivity desirable for a particular disease state.

The starting point for this work was the highly potent, NK3-selective quinoline analog, SB-223412 (Talnetant, **7**). Early SAR on these compounds showed that the quinoline nitrogen is essential for antagonist activity. Furthermore only some very limited substitution was allowed in the pendant 2-phenyl ring, suggesting that this group was accommodated in a tight binding pocket. Conformational analysis of the 2,3,4-trisubstituted quinolines shows that the 2- and 4- substituents must sit orthogonal to the quinoline ring. Thus only two essential conformations are possible (the 4-amido group sits above or below the plane of the quinoline as drawn). It was interesting to note also that only the S enantiomer of the 4-amide substituent was active.

Although many initial docking poses were tried, minimization led to only one especially good docked structure. In this (Fig. 11.8) the quinoline nitrogen was hydrogen bonded to the hydroxyl of tyr_6:51 [31]. This residue is also found in NK2
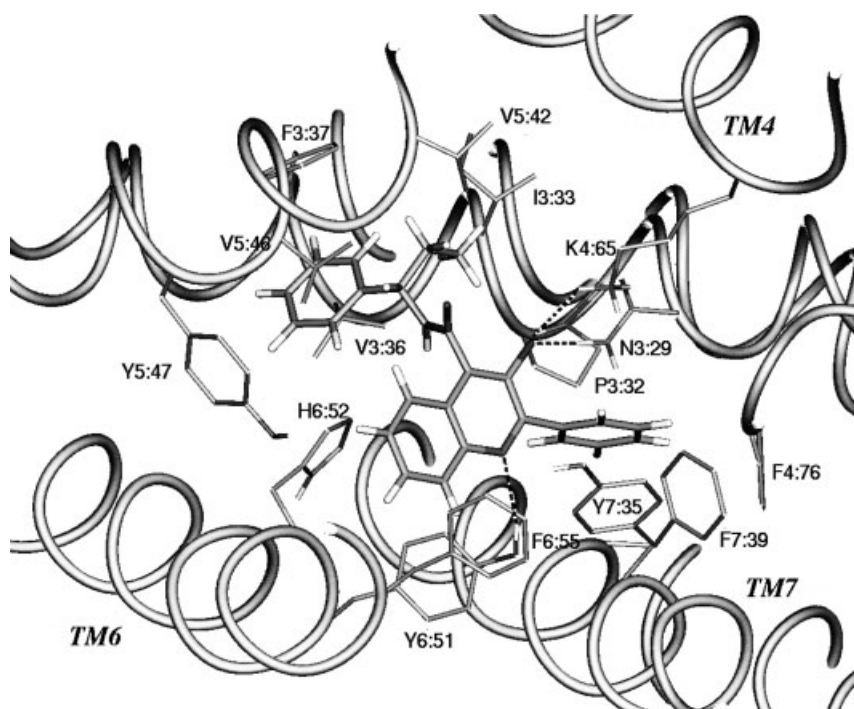
Fig. 11.8 Docking of Talnetant into the NK3 model showing key interactions.

but is a phenylalanine in NK1, in keeping with the series' total lack of binding at the latter receptor. The quinoline ring itself sat in an aromatic pocket formed by tyr_5:47, phe_6:55 and his_6:52. His_6:52 also formed a hydrogen bond with the NH of the ligand amide. As suggested from the SAR, the 2-phenyl ring was contained in a tight hydrophobic pocket bounded by pro_3:32, asn_3:29, tyr_7:35, phe_7:39 and phe_4:76 (in ECL2). All except the last of these residues are conserved between the NK2 and NK3 receptors. Finally a hydrogen bond was found between asn_3:29 and the phenolic group of the ligand. The key differences between NK3 and NK2 are the presence of lys_4:65 at the top of TM4 (threonine in NK2), the identity of several potential binding residues in ECL2 such as leu_4:74 (lysine in NK2) and the nature of the hydrophobic side chains around the 4-amido substituent. Because of the residue conservation it was reasonable to assume that no improvement in NK2 binding would be likely with modifications to either the quinoline ring or its 2-phenyl substituent. However it was thought likely that introduction of a suitably sized protonated amino group instead of the acidic phenol at the 3-position, would allow interaction with the polar residues in the second extracellular loop of NK2 while at the same time improving CNS penetration. Another area of the ligand which was thought worth targeting was the 4-amido substituent where evident differences in the steric nature of the residues of the two receptors could be seen. In this region while

phe_3:37 and ile_3:33 were conserved, val_5:46 was an isoleucine in NK2 and val_3:36 was a methionine.

As predicted from docking, no improvement in selectivity was obtained by introduction of a 3-dimethylaminomethyl substituent in place of the phenol. According to the model, the protonated nitrogen of this compound interacted with the conserved asn_3:29. The model did however suggest that there was room to extend the 3-substituent through the cavity between cys_4:61 and val_5:42 towards ECL2. Here it could interact with lys_4:74 and his_5:39. Docking of the secondary 3-piperazinylmethyl analog in the NK2 model suggested that the compound could make an interaction with lys_4:74 through a strong amine-protonated amine hydrogen bond (Fig. 11.9a). Furthermore alkylation of the secondary piperazine nitrogen should increase the affinity at both receptors through hydrophobic interactions with thr_4:65, cys_4:61 and val_5:42. The results in Table 11.1 confirmed this with the N-cyclohexylpiperazine analog (SB-327076, **8**) showing equipotent nanomolar activity at both NK2 and NK3 (Fig. 11.9b). Thus the first goal of achieving a potent mixed NK2/NK3 antagonist from a selective NK3 starting compound had been achieved.

Table 11.1 Variation in NK2, NK3 and μ-opioid activity with the substituent on the 3-methylene position of Talnetant.

Substituent	Piperazine	N-*isopropyl*-piperazine	N-cyclohexyl-piperazine (9)	Morpholino-piperidine (10)
hNK3 IC50 (nM)	4.7	1.8	1.2	0.8
hNK2 IC50 (nM)	22.2	5.7	1.6	0.8
Ratio	4.7	3.2	1.3	1.0
μ-opioid IC50 (nM)	203.1	103.6	15.8	>2500

As will be seen later, it is always important in ligand docking to consider the electronic structure of the compounds involved. The piperazine analogs can in theory, exist in two different protonated forms. *Ab initio* quantum calculations (HF/6−31G*) on the two forms of SB-327076 showed that protonation at the cyclohexyl substituted nitrogen was favored by 4.3 kcal/mole in terms of absolute energy. However a better estimate of the relative basicity of multiple protonation sites can be obtained by calculating the relative positive electrostatic potentials around each protonated isomer, the theory being that the more basic the nitrogen, the more polarized the corresponding NH bond will be in its protonated state, and hence the greater the positive potential. These calculations showed that the difference in potential between the two forms of SB-327076 was now only 0.37 kcal/mole and therefore that protonation could likely occur at either nitrogen. Docking of the benzylic protonated form showed that the NH formed a strong hydrogen bond with the amide of asn_3:29, allowing lys_4:74 to make a salt bridge with asp_4:69. This pose was more favorable energetically than the alternative
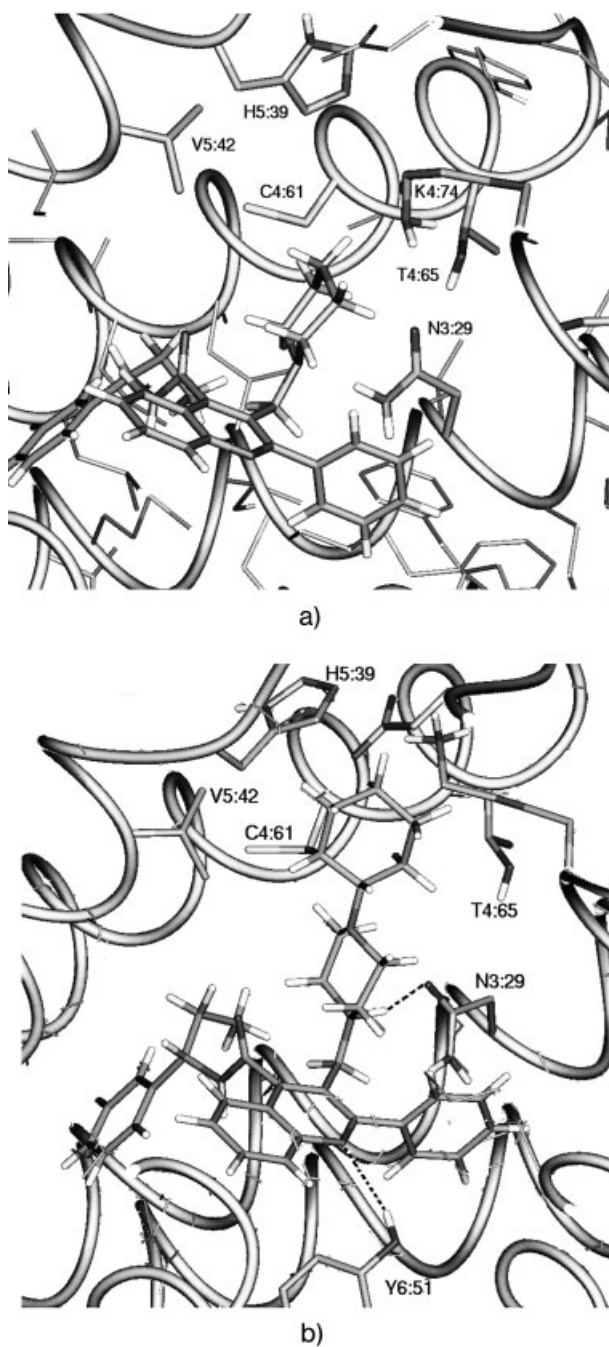
Fig. 11.9 Key residues involved in the binding of (a) the secondary 3-methylpiperazinyl analog of Talnetant and (b) SB-327076 to the NK2 receptor.

form protonated at the remote nitrogen and forming a hydrogen bond with lys_4:74 (Fig. 11.9 b).

It can be seen from Table 11.1 that as the NK2/NK3 potency of the series increased, an undesired side effect started to emerge, namely an increasing antagonist activity at the μ-opioid receptor. The opioid receptors all contain an aspartate residue in TM3 at the same position as the conserved binding aspartate of the aminergic receptors. SDM studies have shown that opioid ligands, almost all of which contain a protonated nitrogen, bind to this aspartate. Docking of SB-327076 essentially led to a single pose in which the conformation of the ligand was unchanged with respect to the NK docked orientations. However two key differences were that (1) to have an interaction with the TM3 aspartate, protonation must occur at the cyclohexyl substituted nitrogen and not at the benzylic position, and (2) the compound docked in a completely inverted position relative to the NK orientation (Fig. 11.10).

The rigid nature of the docking around the cyclohexylpiperazine (in the opioid receptor) suggested that moving the nitrogen by one atom into the cyclohexyl ring would prevent interaction with the TM3 aspartate. NK affinity however should not be affected because protonation at the benzylic nitrogen is preferred in these receptors. The compound (SB-371059, **9**) was synthesized and it was found that it now showed 1000 × selectivity for NK3 (and NK2) over the μ-opioid. The second piperidine ring was now at the top of the 3-substituent binding pocket. It was
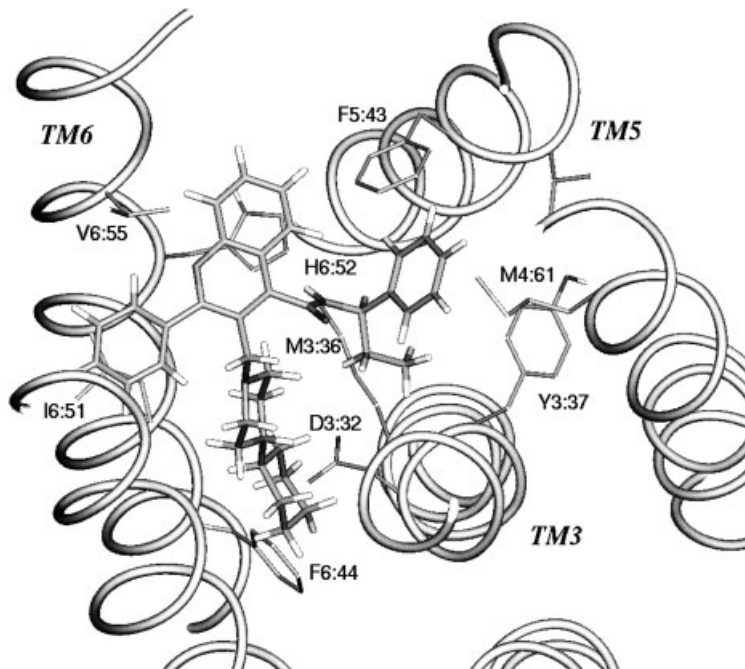
Fig. 11.10 SB-327076 docked into the μ-opioid receptor model with the protonated nitrogen interacting with asp_3:32.

therefore reasonable to assume that suitable substitution would pick up additional hydrogen-bond interactions with the conserved TM4 ser_4:64 and his_5:39. Changing the second ring to a morpholine did achieve this and led to an equipotent sub-nanomolar compound (SB-400238, **10**) with exquisite selectivity over the μ-opioid receptor (Table 11.1).

While the goal of obtaining a potent mixed NK2/NK3 antagonist had been reached in this series, there was still a need to develop a completely selective NK2 analog. An increase in NK2 affinity had been achieved by variation of the 3-substituent as seen above. This however had no effect on the potency of the compounds at NK3 and therefore no further changes in the series at this position were considered. Instead attention turned to the nature of the 4-amido substituent which the docking studies had suggested bound into a hydrophobic pocket where differences did exist between NK2 and NK3. Using docking to guide the synthesis a series of alterations were made to the 4-substituent. This resulted in the synthesis of the most selective NK2 antagonist in the series (SB-415707, **11**).

11.7
Homology (Fifth Generation) Models of 7TM Receptors

Excitement in the 7TM modeling community reached a peak in 2000 with the publication of the crystal structure of bovine rhodopsin [32]. For the first time it was possible to use direct homology methods to construct a new (fifth) generation of 7TM models. Confidence in these models was increased because rhodopsin itself contains the aforementioned conserved residues in each TM helix which play a key functional or structural role. Alignment of other 7TM sequences is therefore made simple in the TM helical regions except in the occasional receptor where one or more of these conserved motifs are missing.

An automated program has been written to build these models directly from the rhodopsin crystal structure and a database of alignments of the TM regions of all known 7TM receptors. The same loop building strategy described earlier is used here but a 3D database of all common loop constructs has been generated to make the initial starting model easier to build. The program generates a complete CHARMm input file for the structure which can then be run manually by the user.

11.8
"Ligand-based" Design of 7TM Receptor Compounds

As mentioned at the beginning one common method of lead generation is by database searching using a pharmacophore. In the ideal world of "structure-based" drug design one would typically rely on the availability of an X-ray structure of the target, preferably with a number of diverse ligands bound into it. One well-known example of this is the HIV protease where many research groups have solved

ligand-bound crystal structures and used these in rational drug design [33]. Unfortunately in most cases, including the 7TM receptors of interest to pharmaceutical companies, there is a lack of knowledge concerning the receptor/enzyme structure, and this has led to a number of other approaches, based on properties of the known ligands themselves – hence the term "Ligand Based Design". One commonly used concept in this is that molecules which bind to the same site must exhibit some similarity to each other. The development of the concept of "Molecular Similarity" was the subject of much research in the late 1980s/early 1990s [34]. This could be based simply on shape but it was more common to look at some electronic property such as electrostatic potential or field or other empirical properties such as hydrophobicity. The gnomonic projection technique which we and others developed led to some surprising ideas on how differing molecules could be overlaid in their receptor-bound conformations [35–37]. It was often useful to combine several properties in the similarity score and this led to useful programs such as the SEAL method [38]. When overlaying a set of diverse molecules, rather than using calculated properties, it is often faster to use empirical factors such as acidic or basic sites, aromatic ring centroids, hydrophobic groups and hydrogen-bond donors and acceptors. Conformational flexibility can also be accounted for in generating these overlays. The overlay, once generated, is generally known as a "pharmacophore".

Pharmacophores can offer several potential advantages, for example, they often present a rationale for ligand SAR. Their main advantage and use however is that they can be used for database searching, which may in turn give rise to novel lead series. Pharmacophores can be generated manually but it is more common to use one of the available software packages such as Catalyst [39]. These programs can also be used for the database searching stage. Over 100 pharmacophores have been published for 7TM receptor targets. One successful application from our group is outlined below.

Once again we were interested in discovering novel compounds as 5-HT_{2C} antagonist leads. Pharmacophore modeling was therefore extensively applied as part of an integrated medicinal and computational chemistry strategy aimed at identifying such novel antagonists. For this purpose, a small set of representative structures was selected from the wide "in-house" collection of 5-HT_{2C} antagonists to span the available chemical diversity. All compounds included in the study had a high affinity in a 5-HT_{2C} binding assay ($pKi > 8.5$), and were at least 100-fold selective over 5-$HT_{2A/2B}$. In addition, they were assumed to bind at the 5-HT_{2C} receptors with a common binding mode which, according to the models described earlier, consisted of hydrogen bonds to the urea/amide carbonyl and a salt bridge between the protonated nitrogen and the TM3 aspartate. The compounds used to build the pharmacophore are shown in Fig. 11.11.

In this example all the pharmacophore modeling work was performed with the Catalyst program. Ligand conformational searches were performed with an energy window of 10 kcal using a maximum of 250 conformations. The common Catalyst pharmacophore features (H-bond acceptor, positive ionizable groups, aromatic rings, hydrophobic groups) were aligned using the HipHop algorithm within the
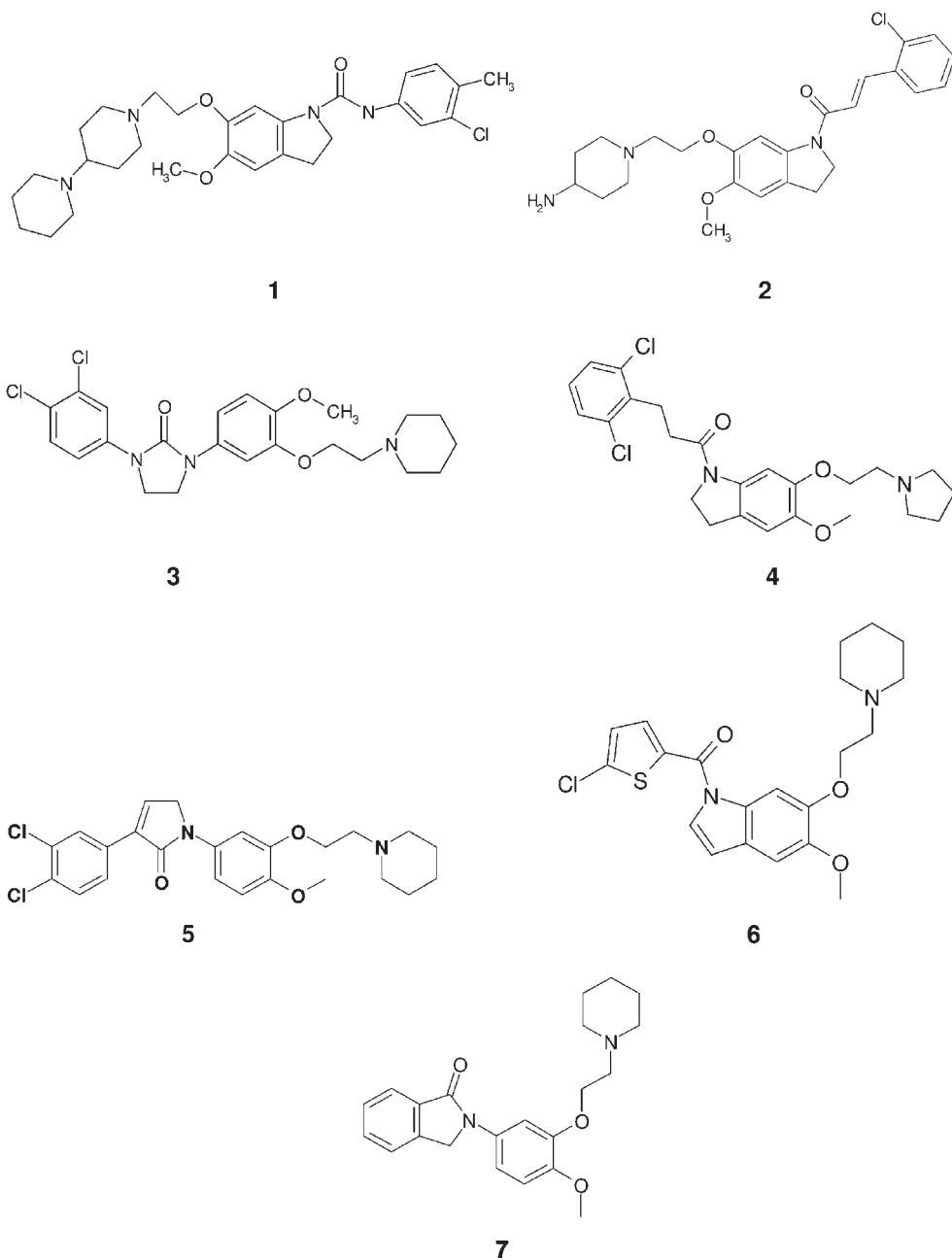
Fig. 11.11 Compounds **1–6** are different 5-HT$_{2C}$ antagonists used in the construction of the pharmacophore shown in Fig. 11.3**c**. Compound **7** was a novel lead generated from a search of this pharmacophore which had good activity on 5-HT$_{2C}$ (pKi > 7) and selectivity towards 5-HT$_{2A/2B}$.

program. From the different top scoring solutions generated by the program, the most satisfactory models were chosen according to quality of ligand conformations, RMS deviation of the ligand conformations from the pharmacophore features and common volumes. In Fig. 11.6 b the best pharmacophore model obtained is shown superimposed on some representative compounds included in the study.

As can be seen from this figure (Fig. 11.6), the pharmacophore model was made up of a positive ionizable group (red sphere), which is mapped by the basic nitrogen on the piperidine side chain of compound **3**, a H-bond acceptor (green spheres) mapped by the carbonyl oxygen on the cyclic urea moiety of compound **3**, and three hydrophobic groups (cyan spheres). This pharmacophore was supported by the SAR derived from "in-house" data which clearly demonstrated the key roles played by the protonated nitrogen moiety and the hydrogen-bond acceptor functionality of the urea or amide carbonyl, both of which were necessary for high 5-HT_{2C} affinity. A second function of the hydrogen-bond acceptor, mapped by the ether oxygen on the piperidine side chain of compound **3**, was instead discarded as it was deemed irrelevant for high 5-HT_{2C} affinity.

The best 5-HT_{2C} pharmacophore solutions were used as a 3D search query against our Catalyst corporate database. About 19,000 structures were retrieved which fitted the pharmacophores and hence were, in theory, endowed with a probability of exhibiting good 5-HT_{2C} activity. This list of compounds was subsequently pruned using filters developed "in house" for the removal of reactive and undesirable molecules, as well as of those predicted to have inappropriate physico-chemical properties for a CNS drug. This reduced list of structures was finally prioritized according to their fit to the pharmacophore. From it, the 246 top scoring compounds were ordered and tested against 5-HT_{2C}. The results were extremely encouraging in that ~32% were found to show 5-HT_{2C} antagonist binding in the sub-micromolar range. In addition, ~5% out of them were also selective against 5-HT_{2A} and 5-HT_{2B} receptor subtypes.

Many of these active compounds belonged to structural classes different from those used to build the pharmacophore (Tanimoto index < 0.57), thus confirming the general validity of the method. One template in particular (Fig. 11.11, compound **7**) was especially potent in the 5-HT_{2C} assay and was also selective towards 5-HT_{2A} and 5-HT_{2B} [40].

11.8.1
Pharmacophores Often do NOT Work

In the last decade, much evidence has emerged suggesting a pathological role for the chemokine polypeptide, MCP-1 and its receptor, CCR2, in diseases involving chronic inflammation, in particular atherosclerosis [41]. Thus, MCP-1 has been shown to be highly expressed in human and mouse atherosclerotic lesions by *in situ* hybridization or immunostaining. Furthermore, knockout mice in which the genes for MCP-1 or CCR2 have been silenced, show a substantial reduction in atherosclerotic lesions, whilst overexpression of mouse MCP-1 in mice was found

to exacerbate the disease. These data suggest that lowering MCP-1 levels, or blocking CCR2, in man may have useful therapeutic effects in treating chronic inflammatory conditions such as atherosclerosis and/or arthritis and considerable effort has been expended by the pharmaceutical industry towards the discovery and development of small molecular weight CCR2 antagonists.

Four chemically distinct series of CCR2 antagonists have been reported but all have potential problems as viable drug leads. The cinnamide **13** (SB-282241) is a potent functional antagonist of CCR2 [42], but suffers from lack of selectivity over 5-HT receptors. The spiropiperidine **14** (RS-504393) [43] has equal affinity at the $á_{1B}$ adrenoceptor, and also has lower functional potency than would be predicted from its binding affinity. Several publications from Takeda have highlighted the CCR5 receptor antagonist **16** (TAK-779) [44], which also possesses significant activity at the CCR2 receptor. Acceptable bioavailability of this quaternary salt may be an issue, so clearly the profile of the corresponding free base **15** is also of interest. Finally a series of amide derivatives have been reported from Teijin as possessing CCR2 antagonist activity, with **17** being identified as a lead compound [45]. Both compounds **14** and **17** are reported to have poor affinities at rodent CCR2 receptors making biological evaluation of these compounds difficult.

In addition to improving the activity and profile of SB-282241, we were interested in generating a pharmacophore, based on the above compounds, which could be used to search for new lead compounds for this target. Assisted by the fact that all compounds possessed a positively charged center which was essential for activity, both manual and computer generated (Catalyst) overlays were derived which were essentially the same (Fig. 11.6**c**). Searching the database against this pharmacophore however only returned compounds related to the original series and no new leads were found. The probable reasons for this failure lie in inherent problems with the classical pharmacophore approach. To start, there is an assumption that all ligands bind in exactly the same site. Therefore the rigid overlap of key features such as basic nitrogens or aromatic rings is overemphasized. There is also an emphasis on the anisotropic nature of certain molecular interactions such as the assumption of fairly rigid directionality with hydrogen bonds. All these assumptions arise from the approximation that the protein remains essentially rigid. Potentially therefore a better approach would be to dock the various ligands into the receptor, taking into account the possibility of multiple binding sites with full protein flexibility, and making use of other available data such as ligand SAR, NMR or SDM results. The overlay of the docked positions should therefore produce a better pharmacophore on which to carry out database searches. Such a combined modeling/SDM approach was used with the CCR2 receptor and the various compounds detailed above [46].

11.9
Refinement of 7TM Pharmacophores Using Current Receptor Models

A previous publication had highlighted the importance of glu_7:39 in the binding of the Roche compound, RS-504393 and as this residue is completely conserved in chemokine receptors, it was initially assumed that the positively charged groups of all compounds bound at this position. Initial docking experiments suggested the involvement of a number of other residues and these were all the subject of SDM studies. In addition to the E7:39 → Q mutation which had a large detrimental effect on binding with all the compounds tested, other potential hydrogen-bonding residues in TM7 were also mutated. Thus none of the mutations D7:32 → A, Q7:36 → A and T7:38 → A had any effect on ligand affinity. In contrast (Fig. 11.12), the mutation T7:40 → A caused a significant drop in binding, with the Roche and Teijin compounds being most affected. Some TM3 aromatic residues were also mutated; here both Y3:32 → A and H3:33 → A showed significant decreased affinity for all compounds. Interestingly however the H3:33 → F mutation showed a small but significant increase for the Takeda analogs and a small decrease for the others. The Takeda compounds were synthesized as CCR5 antagonists and this histidine residue is a phenylalanine in the CCR5 sequence. An extremely interesting finding was observed with the Y1:39 → F mutation in TM1. Here a significant increase in affinity was observed for the Roche and SB compounds whereas a decrease was seen for the others. In fact in several analogs of the SB series, an increase of up to 1000-fold was observed (data not shown). This could only
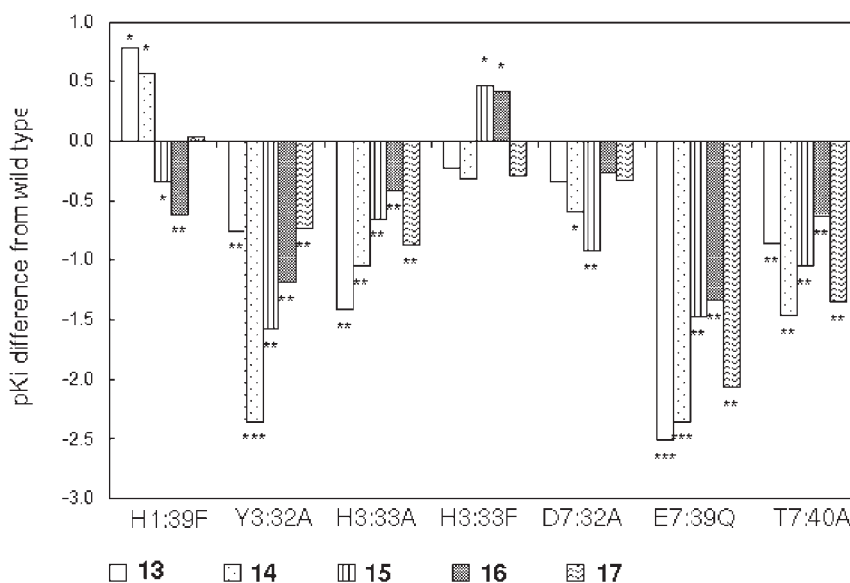
Fig. 11.12 Results from SDM studies on the CCR2 receptor with compounds **13–17**. (Courtesy by J. Med. Chem. (2000), 43, 1123–1134, Copyright of ACS 2000).
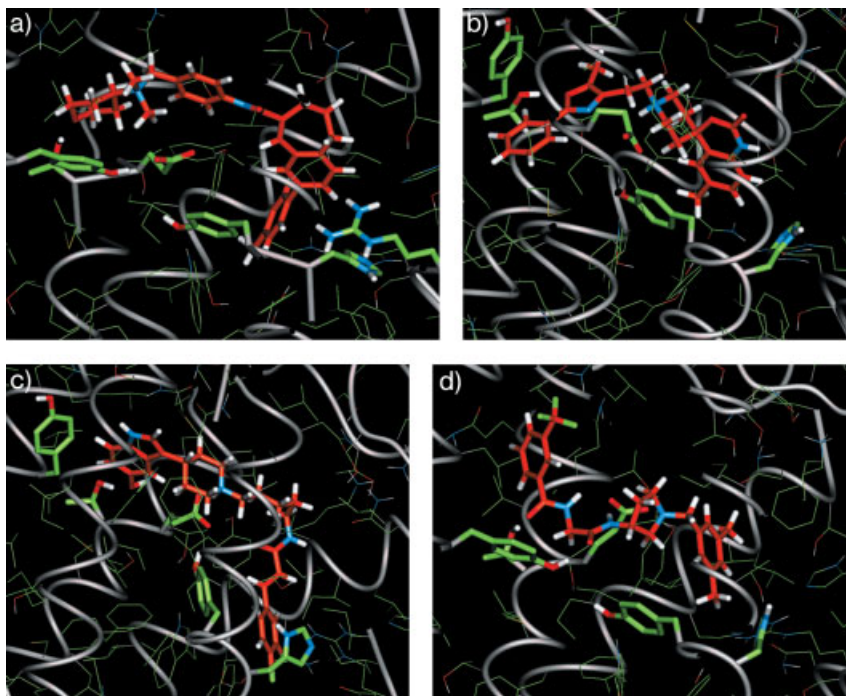
Fig. 11.13 (a) TAK-779 bound to CCR2. Primary binding of quaternary nitrogen to glu_7:39 with an extra polar interaction with the tyr_1:39 hydroxyl. Tyr_1:39 itself has a hydrophobic interaction with the tetrahydropyran ring and thr_7:40 H-bonds to the ring oxygen. The closer proximity of the tertiary nitrogen to glu_7:39 in **15** indicates that the THP oxygen is closer to thr_7:40. Biaryl is in a – sandwich with tyr_3:32 and tyr_6:51 and has only a small interaction with his_3:33. There is increased aromatic interaction with the H3:33 F mutant (see CCR5). (b) RS-504393 binding to CCR2. Primary interaction is with glu_7:39 and an H-bond between oxazole and thr_7:40. Tyr_1:39 flips to the *gauche+* rotamer to accommodate the phenyl ring. The bicyclic ring fits into the aromatic pocket with coplanar

π-stacking to tyr_3:32 and orthogonal interaction with his_3:33. (c) SB-282241 with CCR2. Again primary binding is to glu_7:39 with an H-bond between the 5-hydroxyl and thr_7:40. Tyr_1:39 also flips to the *gauche+* rotamer to accommodate the indole. The ligand's alkyl chain is in an all-staggered conformation. Tyr_3:32 is in a hydrophobic interaction with the alkene while the cinnamide ring is in coplanar π-stacking to his_3:33. (d) Teijin analog with CCR2. Here glu_7:39 binds to the pyrollidine nitrogen and to a backbone NH. Tyr_1:39 and thr_7:40 bind to backbone carbonyls. There are generally weaker hydrophobic interactions between the 3,5-dimethylphenyl ring and tyr_3:32 and his_3:33.
(Courtesy by J. Med. Chem. (2003), 43, 4070–4086, Copyright of ACS 2003).

be explained by the fact that in the unliganded model of the receptor, the hydroxyl of tyr_1:39 formed a strong hydrogen bond with glu_7:39, locking the former in a *trans* rotameric state (seen also in Fig. 11.13 a). Docking the Takeda and Teijin compounds suggested that they could readily be accommodated by this orientation (Fig. 11.13 a and d). On the other hand the Roche and SB analogs (Fig. 11.13 b and c) could only be docked when tyr_1:39 adopted the less favored *gauche+* rotamer. This requires breaking the hydrogen bond and thus is made easier when the hydroxyl is removed, i.e. by mutating Y1:39 → F.

The docking of the five compounds together with a summary of the interactions is shown in Fig. 11.13. The overlap of the docked poses (Fig. 11.6 d) now shows significant differences from the original pharmacophore. In particular the nitrogens of the positive center are separated by up to 5.4 Å. This large spread is easily explained by the fact that the glutamate binding residue adopts two different rotameric states depending on the compound, a fact that is not readily handled by the standard pharmacophore approach. There was still significant overlap of the hydrophobic regions but no common hydrogen-bond acceptor site was observed. Searching the database with this model did in fact yield some novel leads which are being followed up. This alternative pharmacophore generation would appear to offer advantages where traditional methods have failed. It does of course take much longer to generate and can be dependent on additional experimental data.

11.10
Optimizing Properties of the CCR2 Antagonists

SB-282241 was originally developed from the indolepiperidinyl biphenyl analog (**12**) which came from a series developed as 5-HT antagonists. In doing so the CCR2 affinity had been improved from 5.3 μM for **12** to 50 nM for **13** [42]. The compound was not surprisingly, still non-selective over various 5-HT receptor subtypes and more worryingly showed no activity against rodent receptors. This is of particular concern because *in-vivo* animal tests are necessary for the progression of any potential drug molecule.

To improve its selectivity over aminergic 7TM receptors SB-282241 was docked into the 5-HT$_{2C}$ model (Fig. 11.14). The indolepiperidine fits into the classic pocket between asp_3:32 and the aromatic residues in TM5 and TM6. The back face of the piperidine is covered by asn_6:55 on TM6. In contrast, in the CCR2 model (Fig. 11.15) the back face of the piperidine sits in a hydrophobic cavity between phe_3:28 and val_4:74 in ECL2 adjacent to the disulfide bond. This suggested that a hydrophobic substituent such as an ethylene bridge could ideally fit into this cavity. The docking of the tropane analog (**18**; Fig. 11.15) and its subsequent synthesis did indeed lead to a marked increase in selectivity over all 5-HT subtypes and in addition, this compound was the first analog to show activity with the mouse receptor.

A common method of increasing affinity within a series of analogs, is by introduction of rigidity (i.e. decrease the entropy) in a flexible part of the molecule.
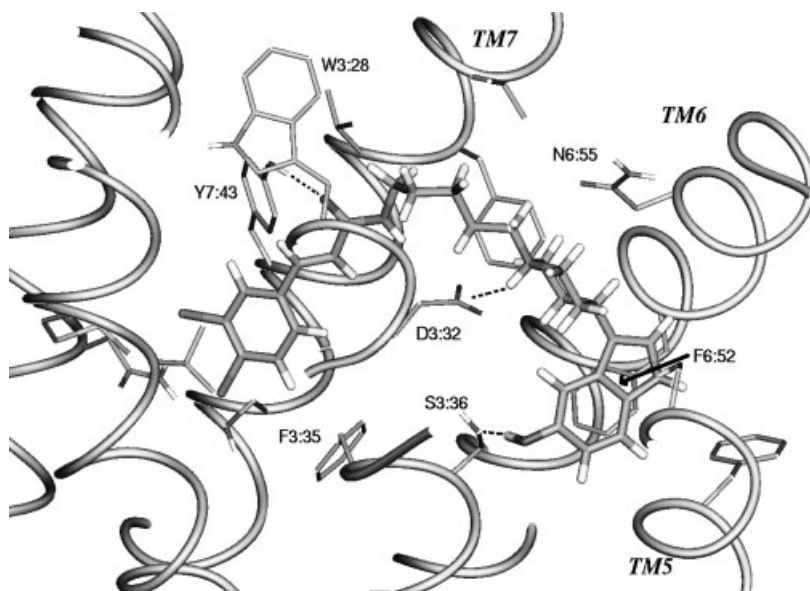
Fig. 11.14 SB-282241 docked into the 5-HT$_{2C}$ model with the primary interaction to asp_3:32. Asn_6:55 forms an unfavorable interaction with the back face of the piperidine ring.
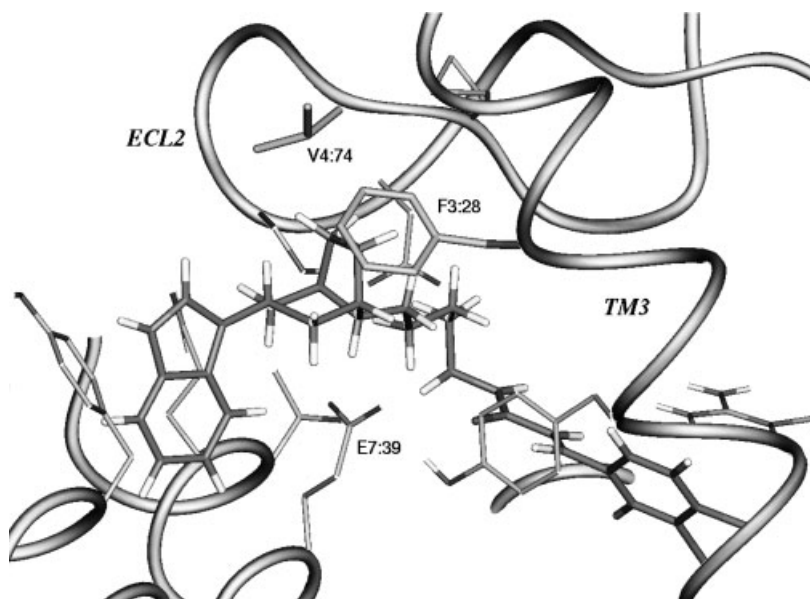

Fig. 11.15 The tropane analog, **18**, docked into CCR2. Now the ethylene bridge forms favorable interactions with val_4:74 and phe_3:28 but an even more unfavorable interaction with asn_6:55 of 5-HT$_{2C}$.
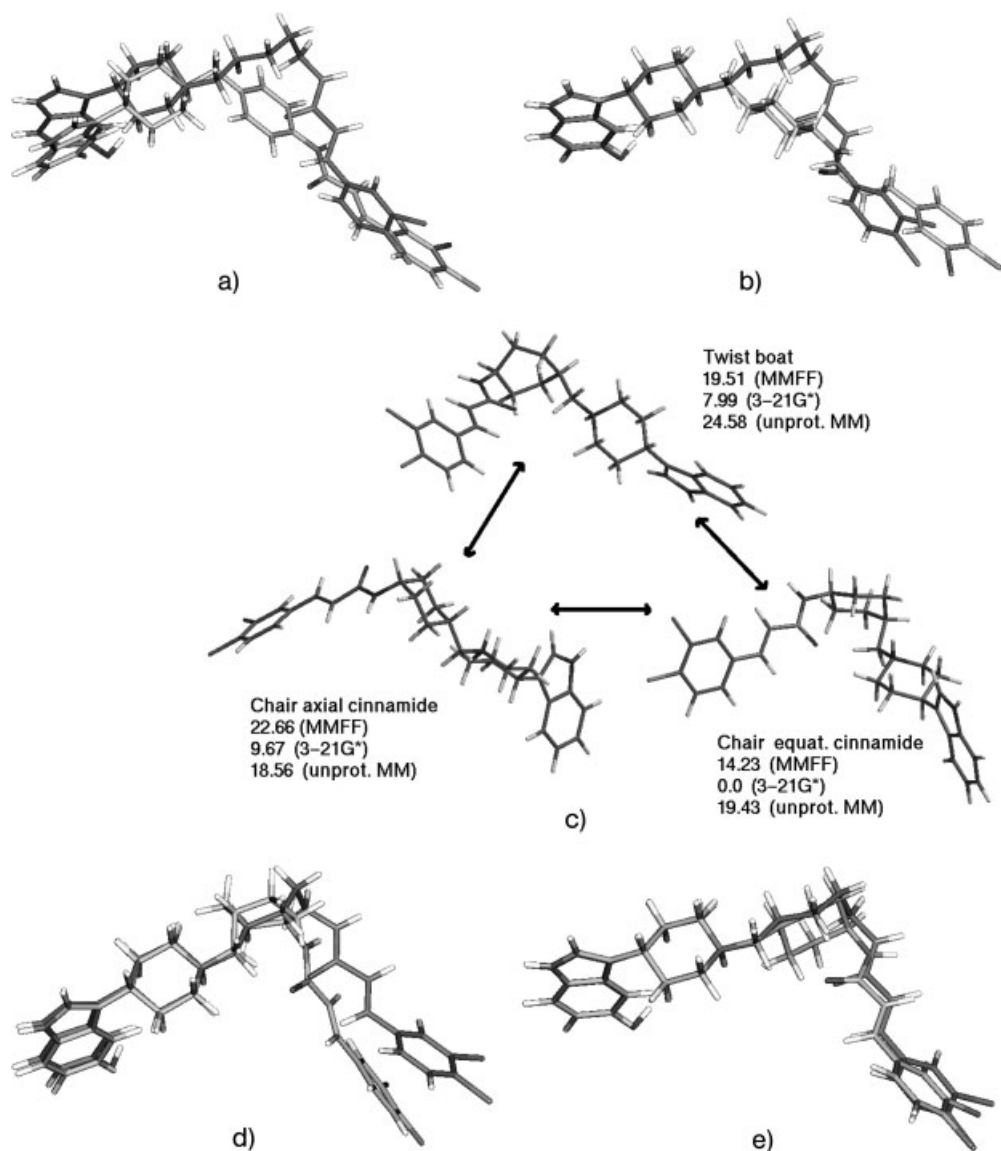
Fig. 11.16 (a–e) CCR2 receptor-bound conformation of SB-282241 (dark) optimally overlapped with its (a) 1,4-phenyl, (b) *trans*-1,4-cyclohexyl, (c) axial cinnamide *cis*-1,4-cyclohexyl and (d) equatorial cinnamide *cis*-1,4-cyclohexyl linker analogs. Only (d) overlays the bound conformation well and has the most favorable activity. (c) *cis*-1,4-cyclohexyl linker analog in three conformational forms. With the protonated form the axial cinnamide conformation has the lowest energy but is stabilized by a strong electrostatic interaction between the amide carbonyl and the protonated nitrogen. In the neutral form the equatorial cinnamide conformation is favored.

Within the SB series, the five-carbon linker was an obvious target for this approach, but although the pentyl chain was in an all-staggered conformation, it was not immediately obvious how a ring linker could emulate this. Shown in Fig. 11.16 is the bound conformation of SB-282241 optimally overlaid with 1,4-phenyl (a) and *trans*-cyclohexyl (b) linkers. Neither shows any correspondence of the amide groups and the dichlorophenyl ring in each is buried much deeper in its binding pocket. The 1,4- and 1,3-phenyl linked analogs were completely inactive and the *trans*-cyclohexyl analog was 3-fold less active in the CCR2 binding assay. The conformation of these rings is obvious but the corresponding *cis* 1,4-cyclohexyl analog needed more careful analysis. Chair forms of this ring must have one axial and one equatorial substituent. They can interchange between these through a number of low energy twist-boat conformations. With the protonated (presumably the bound) form of the molecule the equatorial cinnamide substituent is favored by *ab initio* calculations. However the protonated conformations are stabilized by a strong intramolecular interaction between the positively charged hydrogen and the amide carbonyl and it is therefore more appropriate to use the unprotonated forms (Fig. 11.16 c); in this case, the axial cinnamide isomer is favored. In Fig. 11.16e, it can be seen that there is a perfect overlay between this isomer and the bound conformation of SB-282241 which cannot be found with the other conformers (Fig. 11.16 d). Synthesis and testing of the *cis* 1,4-cyclohexyl-methyl analog compound showed that it had double the potency of the straight chain analog [47]. Incorporation of all these features, i.e. the *cis* cyclohexyl ring linker, the tropane ring and the 5-hydroxy substituent which forms an additional hydrogen bond with thr_7:40, led to a compound, SB-380732 (**19**) with very good activity (20 nM in human and mouse chemotaxis assays) and excellent selectivity over other chemokine and aminergic receptors. The compound *per se* had some clearance issues but these were easily remedied by replacing the 5-hydroxy substituent in the indole ring with a 5-methylsulphonamide.

11.11
Some General Ligand Considerations When Docking

The above examples have detailed how 7TM receptor models have improved over the last 15 years and how as a result, structure-based design has improved in expectations. The models are certainly at a stage where *de novo* design should be possible if the computer programs were accurate enough. Until this happens however it is still necessary to start from a suitable lead either from high throughput screening of compound libraries/collections or indirectly through initial *in silico* selection and subsequent evaluation. Structure-based design is still best suited to lead optimization through improvements in selectivity and activity. The starting point for this is an accurate docking of the lead ligands which is best done manually with full consideration of the compound's properties. Stereochemistry is one obvious starting point. A compound may contain a chiral center(s) and the active enantiomer or diastereomer may be unknown. In this case it will be necessary to
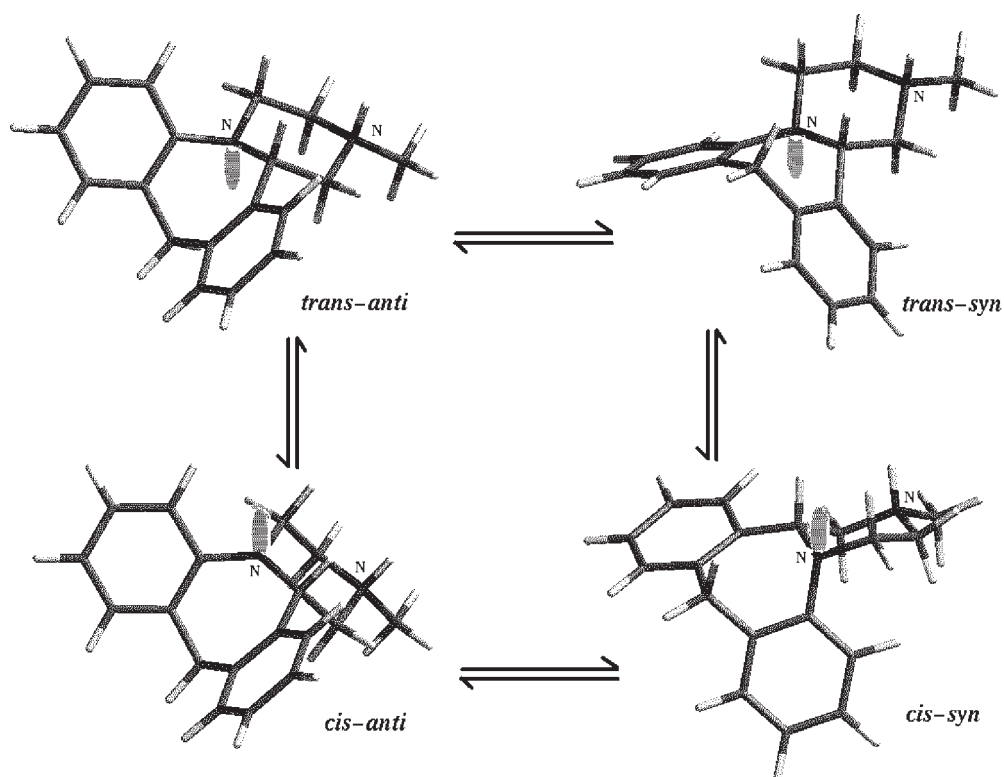
Fig. 11.17 Mianserin in its four ring conformations.

dock all possible isomers, a fact which should be obvious but is often ignored. The same applies to geometric isomers. The stereochemistry of olefins is usually known but the same is not always obvious for enamines, enol ethers or non-symmetric N,N-disubstituted amides. Here it is necessary to perform careful geometry optimization of all possible isomers using molecular mechanics and/or quantum methods. It may be found that some isomers are too high in energy or that only one may fit into its receptor model. Electronic factors should always be taken into account. The compound may exist as different tautomers or have several potential ionization sites with differing pK_as. Here careful quantum calculations should be made to decide on which isomer(s) to dock. It should be noted that in most cases, the generation of 3D structures from, e.g. SMILES strings will not take these properties into account. Finally a word should be said about conformation. It is evident that low energy states should be docked although not necessarily only the lowest energy form. Conformational analysis is not confined to rotation around flexible bonds or simple ring flipping but also may involve more subtle properties such as nitrogen inversion. An interesting example was recently described where explanations were sought for the observed selectivity of various compounds for the human and mouse 5-HT$_6$ receptors [48]. The antidepressant, mianserin (**20**)
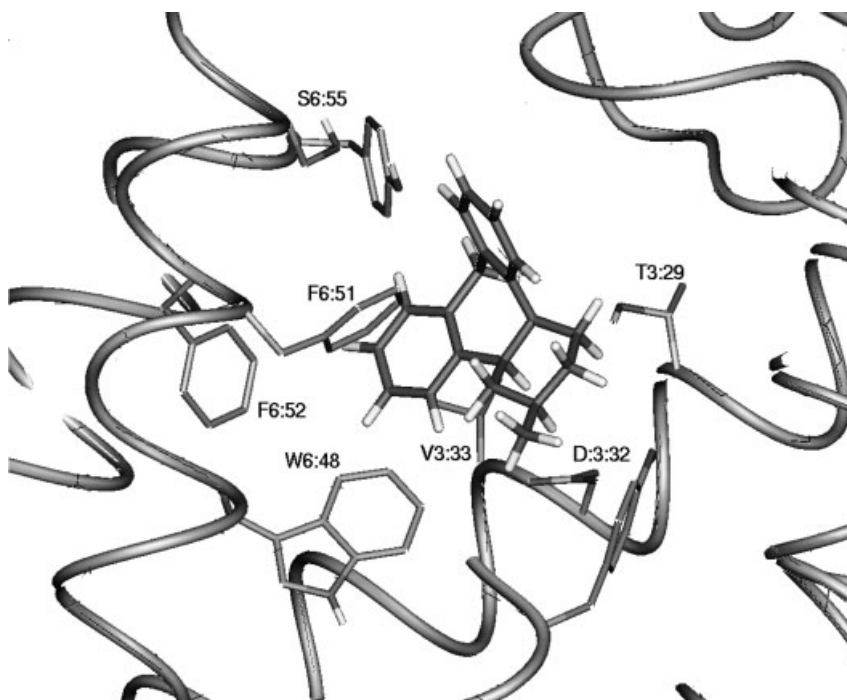
Fig. 11.18 Cis-syn mianserin docked into the mouse 5-HT6 receptor model showing its primary binding with asp_3:32 and other π–π interactions with phe_6:51 and phe_6:52.

was shown to bind selectively to the mouse receptor, which differs from the human by only four residues in the potential binding site. Careful conformational analysis using a combination of molecular mechanics, semi-empirical and *ab initio* quantum methods showed that the tetracyclic ring system exists in four possible forms involving a flip of the azepine ring making the bridgehead hydrogen *syn* or *anti* to a hydrogen of the azepine methylene, and an inversion of the nitrogen placing its lone pair *cis* or *trans* to the bridgehead hydrogen. In mianserin both calculations and experiment show that the *trans-anti* conformer is preferred. The *cis-anti* conformer can clearly be ruled out not only because of its high energy but because under normal optimization conditions, it inverts to an alternative. The two other conformations, i.e. the *cis-syn* and the *trans-syn* are intermediate in energy (Fig. 11.17).

Docking of mianserin in the *trans-anti* conformation to the mouse receptor model predicted that, in addition to the usual salt bridge between the piperazine nitrogen and asp_3:32, a π–π stacking occurred between phe_6:51 and one of the aromatic rings of the ligand, as did an orthogonal aromatic–aromatic interaction between phe_6:52 and the other ring. No other significant interactions were observed with this conformer. With the *cis-syn* conformer however, phe_6:51 formed the more favorable orthogonal interaction with both phenyl rings of the ligand.

Other favorable interactions were found between phe_6:52, phe_7:35, thr_3:29 and val_3:33. Thus the overall binding mode of mianserin in the mouse favors the *cis-syn* conformation (Fig. 11.18). SDM studies showed that only the S6:55 → N mutant (one of the four residues which differ in the mouse and human receptors) decreased the binding to the same level as the human wild-type. The models suggested that there was a large steric clash between the asparagine and one of the phenyl rings of the ligand. Based on these conformations, two pentacyclic analogs were designed and synthesized with the aim of locking the conformation of the ring system in either the *cis-syn* (BRL-34849, **21**) or the *trans-anti* (BRL-34969, **22**) states. These conformations were predicted by *ab initio* energy calculations and subsequently confirmed experimentally. As with mianserin BRL-34849 shows a selectivity of about 10-fold over the human receptor and shows the same decrease with the S290N mutant. BRL-34969 however shows no selectivity for either wild-type or mutant receptors, thus giving experimental proof that the *cis-syn* conformation is indeed the bound conformation.

11.12
What the Future Holds

Nowadays models of the antagonist state of Family A receptors are, in many cases of sufficient quality to be useful in structure-based drug design. This is especially true when docking hypotheses can be tested by SDM experiments. Two areas of current research promise to be of similar use in the future. One is in the development of models of the agonist state of Family A sequences. There is a wealth of experimental work suggesting specific conformational changes between antagonist and agonist states and many groups are now using this information to generate working 7TM agonist models [17]. A second field of research is in modeling Family B and C receptors. Although there is little perceived sequence identity between these and the Family A TM domains, models equivalent to the aforementioned "third generation" structures are now becoming available. Furthermore much of the known binding of the peptide (Family B) and small ligand (Family C) agonists occurs in the N-terminal domains and the structures of these are now known from crystallography [49] and NMR [50]. The unliganded NMR structure of the Family B CRF receptor NTD may well change upon peptide binding but holo- and apo- forms of the Family C NTD have been solved. These therefore promise to be fruitful areas of design in the future.

11.13
Abbreviations and Nomenclature

The terms GPCR (G Protein-Coupled Receptor) and 7TM (Seven Transmembrane helical) are internationally recognized as descriptors of the superfamily of sequences under discussion in this chapter. Likewise the abbreviation TM (trans-

membrane) is equally recognized and will be used throughout the discussion. Some other standard abbreviations include 2D- and 3D- (two- and three-dimensional), SAR (structure-activity relationship), SDM (site-directed mutagenesis) and ECL (extracellular loop). To uniquely identify residues, the nomenclature described by Ballesteros will be used [1]. Here the first number represents the TM helix and the second number is the position within that helix relative to the most conserved residue which is numbered 50. Residues in the ECL's are numbered from the preceding TM helix. For all residues, the standard three-letter code will be used, except where mutations are discussed. Here the standard one-letter nomenclature will be used, where the first letter is the residue being mutated and the second is the mutant. With the exception of the 5-HT$_{2C}$ pharmacophore, all chemical structures discussed in the chapter are numbered as shown in Fig. 11.1 and will be referred to in the text as **(1)**, **(2)**, *etc.*

References

1 BALLESTEROS, J.A., WEINSTEIN, H. *Methods Neurosci* **1995**; *25*: 366–427.

2 FINDLAY, J., ELIOPOULOS, E. *Trends Pharmacol Sci* **1990**; *11*: 492–499.

3 HUMBLET, C., MIRZADEGAN, T. *Ann Rep Med Chem* **1992***; 27*: 291–300.

4 TRUMPP-KALLMEYER, S., HIBERT, M.F., BRUINVELS, A., HOKLACK, J. *Mol Pharmacol* **1991**; *40*: 8–15.

5 TRUMPP-KALLMEYER, S., HOKLACK, J., BRUINVELS, A., HIBERT, M.F. *J Med Chem* **1992**; *35*: 3448–3462.

6 STRADER, C.D., CANDELORE, M. R., HILL, W. S., SIGAL, I.S., DIXON, R.A.F. *J Biol Chem* **1989**; *264*: 13572–13578.

7 KYTE, J., DOLITTLE, R.F. *J Mol Biol* **1982**; *157*: 105–132.

8 ENGELMAN, D.M., STEITZ, T.A., GOLDMAN, A. *Annu Rev Biophys Biophys Chem* **1986**; *15*: 321–353.

9 VRIEND, G. at http://www.gpcr.org/7tm/atricles/model.html **1995**.

10 EISENBERG, D., McLACHLAN, A.D. *Nature* **1986**, *319*, 199–203.

11 SCHERTLER, G.F.X., VILLA, C., HENDERSON, R. *Nature* (*Lond*) **1993**; *362*: 770–772.

12 HOFLACK, J., TRUMPP-KALLMEYER, S., HIBERT, M. *Trends Pharmacol Sci* **1994**; *15*: 7–9.

13 DONNELLY, D., OVERINGTON, J.P., BLUNDELL, T.L. *Protein Eng* **1994**; *7*: 645–653.

14 DONNELLY, D., COGDELL, R.J. *Protein Eng* **1993**; *6*: 629–635.

15 (a) BLANEY, F.E., TENNANT, M. In *Membrane Protein Models*, J.B.C. Findlay (Ed.). Oxford: Bios Scientific Publishers Ltd., **1996**, 161–176. (b) BLANEY, F.E. In *Genomics, Commercial Opportunities from a Scientific Revolution*, Dixon, G.K., Copping, L.G., Livingstone, D. (Eds). Oxford: Bios Scientific Publishers Ltd., **1997**, 77–104.

16 HERZYK, P., HUBBARD, R.E. *J Mol Biol* **1998**; *281*: 741–754.

17 GOULDSON, P.R., KIDLEY, N.J., BYWATER, R.P., PSAROUDAKIS, G., BROOKS, H.D., DIAZ, C., SHIRE, D., REYNOLDS, C.A. *Proteins: Struct, Funct, Bioinformat* **2004**; *56*: 67–84.

18 "HELGA" ADCOCK, S. PhD thesis, Oxford University, **2002**.

19 FLO available from Thistlesoft Inc., Morris Township, NJ, USA.

20 FORBES, I.T., DABBS, S., DUCKWORTH, D.M., HAM, P., JONES, G.E., KING, F.D., SAUNDERS, D.V., BLANEY, F.E., NAYLOR, C.B. *J Med Chem* **1996**; *39*: 4966–4977.

21 BROMIDGE, S.M., DABBS, S., DAVIES, D.T., DUCKWORTH, D.M., FORBES, I.T., HAM, P., JONES, G.E., KING, F.D., SAUNDERS, D.V., STARR, S., THEWLIS, K.M., WYMAN, P.A., BLANEY, F.E., NAYLOR, C.B., BAILEY, F., BLACKBURN, T.P., HOLLAND, V., KENNETT,

G.A., Riley, G.J., Wood, M.D. *J Med Chem* **1998**, *41*, 1598–1612.

22 Bromidge, S.M., Duckworth, M., Forbes, I.T., Ham, P., King, F.D., Thewlis, K.M., Blaney, F.E., Naylor, C.B., Blackburn, T.P., Kennett, G.A., Wood, M.D., Clarke, S.E. *J Med Chem* **1997**; *40*: 3494–3496.

23 Unger, V.M., Hargrave, P.A., Baldwin, J.M., Schertler, G.F.X. *Nature* (*Lond*) **1997**; *389*: 203–206.

24 Kao, H.T., Adham, N., Olsen, M.A., Weinshank, R.L., Branchek, T.A., Hartig, P.R. *FEBS Lett* **1992**; *307*: 324–328.

25 Bromidge, S.M., Dabbs, S., Davies, D.T., Davies, S., Duckworth, D.M., Forbes, I.T., Gaster, L.M., Ham, P., Jones, G.E., King, F.D., Mulholland, K.R., Saunders, D.V., Wyman, P.A., Blaney, F.E., Clarke, S.E., Blackburn, T.P., Holland, V., Kennett, G.A., Lightowler, S., Middlemiss, D.N., Trail, B., Riley, G.J., Wood, M.D. *J Med Chem* **2000**; *43*: 1123–1134.

26 Blaney, J.M., Crippen, G.M., Dearing, A., Dixon, J.S., Spellmeyer, D.C. DGEOM95. Available from the Quantum Chemistry Program Exchange, Bloomington, IN, USA, **1995**.

27 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. *J Comput Chem* **1983**, *4*, 187–217. CHARMm Version 25.2 Revision 98.0731. Accelrys Inc. San Diego, CA, USA, **1999**.

28 Wilmot, C.M., Thornton, J.M. *Protein Eng* **1990**; *3*: 479–493.

29 (a) Elber, R., Karplus, M. *J Am Chem Soc* **1990**; *112*: 9161–9175. (b) Roitberg, A., Elber, R. *J Chem Phys* **1991**; *95*: 9277–9287.

30 Blaney, F.E., Langmead, C.J., Bridges, A., Evans, N., Herdon, H.J., Jones, D.N.C., Ratcliffe, S.J., Szekeres, P.G. (to be published in 2006).

31 Blaney, F.E., Raveglia, L.F., Artico, M., Cavagnera, S., Dartois, C., Farina, C., Grugni, M., Gagliardi, S., Luttmann, M.A., Martinelli, M., Nadler, G.M.M.G., Parini, C., Petrillo, P., Sarau, H.M., Scheideler, M.A.,

Hay, D.W.P., Giardina, G.A.M. *J Med Chem* **2001**; *44*: 1675–1689.

32 Polczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M., Miyano, M. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

33 Kubinyi, H. *Curr Opin Drug Discov Dev* **1998**; *1*: 4–15.

34 Dean, P.M. (Ed.). *Molecular Similarity in Drug Design*. Glasgow: Blackie Academic & Professional, **1995**.

35 Chau, P.-l., Dean, P.M. *J Mol Graph* **1987**; *5*: 97–100.

36 Blaney, F.E., Finn, P., Phippen, R.W., Wyatt, M.J. *J Mol Graph* **1993**; *11*: 98.

37 Blaney, F.E., Edge, C.M., Phippen, R.W. *J. Mol Graph* **1995**; *13*: 165.

38 Kearsley, S.K., Smith, G.M. *Tetrahedron-Computer Methodol* **1990**; *3*: 615–633.

39 Catalyst v4.7. **2002**. Available from Accelrys Inc. San Diego, CA, USA.

40 Bonanomi, G., Hamprecht, D., Micheli, F., Terreni, S. Preparation of phenyl isoindolones with 5-HT2C receptor activity and uses thereof. PCT Int Appl **2004**. WO 2004089897 A1.

41 (a) Ross, R. *New Engl J Med* **1999**; *340*: 115–126. (b) Reape, T.J., Groot, P.H.E., *Atheroscl* **1999**; *147*: 213–215.

42 Forbes, I.T., Cooper, D.G., Dodds, E.K., Hickey, D.M.B., Ife, R.J., Meeson, M., Berkhout, T., Gohil, J., Moores, K. *Bioorg Med Chem Lett* **2000**; *10*: 1803–1806.

43 Mirzadegan, T., Diehl, F., Ebi, B., Bhakta, S., Polsky, I., McCarley, D., Mulkins, M., Weatherhead, G.S., Lapierre, J-M., Dankwardt, J., Morgans, D., Wilhelm, R., Jarnagin, K. *J Biol Chem* **2000**; *275*: 25562–25571.

44 (a) Shiraishi, M., Kitayoshi, T., Aramaki, Y., Honda, S. WO Patent 99/32468. (b) Baba, M., Nishimura, O., Kanzaki, N., Okamoto, M., Sawada, H., Iizawa, Y., Shiraishi, M., Aramaki, Y., Okonogi, K., Ogawa, Y., Meguro, K., Fujino, M. *Proc Natl Acad Sci USA* **1999**; *96*: 5698–5703. (c) Shiraishi, M., Aramaki, Y., Seto, M., Imoto, H.,

Nishikawa, Y., Kanzaki, N., Okamoto, M., Sawada, H., Nishimura, O., Baba, M., Fujino, M. *J Med Chem* **2000**; *43*: 2049–2063.

45 (a) Shiota, T. WO Patent 99/25686.
(b) Tarby, C.M., Endo, N., Moree, W., Kataioka, K., Ramirez-Weinhouse, M.M., Imai, M., Bradley, E., Saunders, J., Kato, Y., Meyers, P. In Book of Abstracts, *218th National ACS Meeting in New Orleans*, Aug. 22–26, **1999**. Poster Abstract No. 82.

46 Berkhout, T.A. Blaney, F.E., Bridges, A.M., Cooper, D.G., Forbes, I.T., Gribble, A.D., Groot, P.H.E., Hardy, A., Ife, R.J., Kaur, R., Moores, K.E., Shillito, H., Willetts, J, Witherington, J. *J Med Chem* **2003**; *46*: 4070–4086.

47 Witherington, J., Bordas, V., Cooper, D. G., Forbes, I.T., Gribble, A.D., Ife, R.J., Berkhout, T., Gohil, J., Groot, P.H.E. *Bioorg Med Chem Lett* **2001**; *11*: 2177–2180.

48 Hirst, W.D., Abrahamsen, B., Blaney, F.E., Calver, A.R., Aloj, L., Price, G.W., Medhurst, A.D. *Mol Pharmacol* **2003**; *64*: 1295–1308.

49 Kunishima, N., Shimadat, Y., Tsuji, Y., Sato, T., Yamamoto, M., Kumasaka, T., Nakanishi, S., Jingami, H., Morikawa, K. *Nature* **2000**; *407*: 971–977.

50 Grace, C.R. R., Perrin, M.H., DiGruccio, M.R., Miller, C.L., Rivier, J.E., Vale, W.W., Riek, R. *Proc Natl Acad Sci USA.* **2004**; *101*: 12836–12841.

12
Receptor-based Rational Design: Virtual Screening

Didier Rognan

12.1
Introduction

G protein-coupled receptors (GPCRs) have historically been the most investigated target family in the pharmaceutical industry for many conceptual reasons: (i) GPCRs select their endogenous ligands among a uniquely broad chemical space [1] which can in turn serves as a basis for designing molecules able to modulate their activity [2]; (ii) their ubiquitous distribution at the surface of many cells for regulating a wide array of physiological and pathological processes render them particularly attractive targets for therapeutic intervention [3]; (iii) many GPCRs have proven over the last 30 years to be druggable [4] and there is reasonable hope that new receptors from this family that are close enough to well-investigated targets could also be addressed by drug discovery programs.

Until recently, GPCR ligand discovery had followed a limited number of strategies including serendipity [5], ligand-based structure–activity relationships [5], and selective optimization of side activities [6]. Concomitant breakthroughs in genomics [7], structural biology [8], bio- and chemoinformatics [9, 10] have considerably widened the number of putative design strategies. Notably the high-resolution crystal structure of bovine rhodopsin [11], the only GPCR for which fine atomic details are available, has paved the way for structure-based design methods [12] which have been successfully applied to enzymes over the last 10 years. Whether receptor-based design should be applied to GPCR ligands (see also Chapter 11) is still a matter of debate, especially with regard to the desired throughput. The objective of present chapter is to review and summarize recent advances in the receptor-based virtual screening of compound libraries and to identify current hurdles that need to be overcome in order to improve the accuracy of this methodology.

12.2
Structure-based Screening Workflow

Regardless of the target under investigation, any protein-based screening is composed of three steps: library design, screening (usually by docking) and data mining. It should be emphasized that each of these three steps is equally important in the whole process.

12.2.1
Setting Up a Ligand Library

The first step is to collect an electronic library of compounds [13] which can be either physically available (*screening library*) or still pending synthesis (*virtual library*). Two main sources for screening libraries are currently possible: in-house catalogs of compounds historically synthesized by pharmaceutical companies [14], and commercially-available screening collections (Table 12.1) [15]. The first type of library is proprietary while the latter is freely available but most probably less enriched in drug-like compounds. In any case, filtering such libraries [16] is an abso-

Table 12.1 Commercially-available screening collections.

Supplier	Library	Size	Web site
Asinex	Platinium	100,000	http://www.asinex.com
	Gold	230,000	
A-Synthese Biotech	Samples for Screening	16,000	http://www.a-syntese-biotech.dk
Aurora Feinchemie	E-catalog	29,000	http://www.aurora-feinchemie.com
Bionet	screening database	43,000	http://www.keyorganics.ltd.uk
Chembridge	EXPRESS-Pick	330,000	http://chembridge.com
ChemDiv	Combilab	250,000	http://www.chemdiv.com
	Diversity	150,000	
ChemStar	ChemStar Library	60,000	http://www.chemstaronline.com
Enamine	Screening collection	400,000	http://www.enamine.relc.com
InterBioscreen	Synthetic	360,000	http://www.ibscreen.com
	Natural	45,000	
Maybridge	Screening collection	60,000	http://www.maybridge.com
MDD	STOCK	34,000	http://www.worldmolecules.com
Otava	Supplier Stock	73,000	http://www.otava.com.ua
Peakdale	Lead-Like Compounds	16,000	http://www.peakdale.com
Pharmeks	Main	105,000	http://www.pharmeks.com
Specs	Diverse	230,000	http://www.specs.net
Timtec	STOCK	164,000	http://www.timetc.net
TosLabs	Compounds collection	23,000	http://www.tolabs.com
Tripos	Leadquest	80,000	http://www.leadquest.com
Vitas-M	STK	196,000	http://www.vitasmlab.com/
	TULIP	25,000	

lute prerequisite in order to remove undesirable compounds exhibiting chemically reactive moieties [17], scaffold-inherited toxicity [18], poor oral bioavailability [19], and promiscuous binding motifs [20]. The filtering intensity should be dependent on the discovery context. For example, filtering might be less restrictive if screening is aimed at discovering a pharmacological tool, and more demanding if the search is for orally-available lead-like compounds. In addition, extra pre-processing steps are still necessary to remove salts, duplicates, to detect and generate stereoisomers/tautomers on the fly, to predict the most likely ionization state and finally to convert 1-D or 2-D sketches into reliable 3-D conformations. In the precise context of GPCR ligand design, a subset of the screening collection fulfilling either known GPCR privileged structures (see Chapter 8) or pharmacophore queries (see Chapter 9) will be generally screened.

However, what is the diversity encoded in such libraries? Of course, answering this question implies that the concept of molecular diversity [21] has previously been addressed, which is not within the scope of the present chapter. Here, the issue of diversity will be addressed from a medicinal chemistry point of view i.e. how many substructures (scaffolds) amenable to fast library design are present in the library? If it is assumed that shape recognition is a very important event in ligand binding, then enrichment in different scaffolds should be a key factor in the evaluation of the quality of a library. By looking at several commercially-available collections covering over 3 million molecules, it is possible to cluster libraries into three-well defined families (Fig. 12.1).
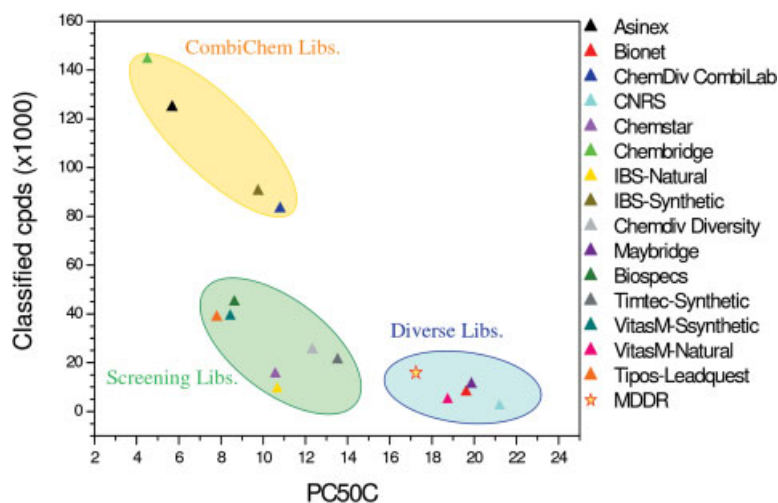
Fig. 12.1 Analysis of commercially-available screening collections. Percentage of scaffolds describing 50 % of drug-like compounds (PC50C metric) of the library. (1) Asinex, (2) Bionet, (3) ChemDiv CombiLab, (4) French National Library, (5) Chemstar, (6) Chembridge, (7) InterBioScreen natural library, (8) InterBioScreen synthetic library, (9) Chemdiv International Diversity collection, (10) Maybridge, (11) Biospecs, (12) Timtec synthetic library, (13) Vitas-M synthetic library, (14) Vitas-M natural compounds, (15) Tripos-Leadquest, (16) MDDR.

"Combichem libraries" (e.g. Asinex, Chembridge, ChemDiv) comprise a large number of drug-like compounds obtained by fast parallel synthesis, but covered by a relatively small percentage of different scaffolds. "Screening libraries" (e.g. Chemstar, Leadquest) are less populated and richer in diverse scaffolds. Last "diverse libraries" (e.g. Bionet, Maybridge) are even less populated but very diverse with respect to their size. It should be noted that most natural compound collections cluster in the latter category. Therefore, the choice of screening collection(s) is a key factor to the success of the screening process. It should be ascertained whether the chemical space covered by a single or several libraries overlaps the target space that is being screened. It is thus strongly advisable to identify non-redundant scaffolds among all available libraries to customize its screening dataset. As most of these collections are quite dynamic entities, it is also advisable to upgrade its stock (approximately every quarter) in order to reflect recent changes in individual collections.

12.2.2
Docking and Scoring

Docking is the computational exercise of predicting both the receptor-bound conformation and the relative orientation of a small molecular weight ligand with respect to its cognate receptor. Usually, a limited conformational space around the putative binding site is sampled. Therefore, a detailed knowledge of the binding site (generally achieved by site-directed mutagenesis) is a very important prerequisite to the screening. It also important to adjust the throughput of the docking to the desired accuracy, modern computing architectures allow a pace of ~1 molecule/min/cpu. Nowadays there are many docking programs [22] available (Table 12.2). They all use the general principle of steric complementarity (e.g. Dock, Fred) or endeavor to maximize the intermolecular interactions (e.g. FlexX, Gold) between the ligand and the receptor.

Up to now, the receptor has been considered as rigid although recently progress has been made to better accommodate at least local flexibility at the binding site

Table 12.2 Main docking programs.

Name	Editor	Web site
AutoDock	Scripps	http://www.scripps.edu/mb/olson/doc/autodock/
Dock	UCSF	http://dock.compbio.ucsf.edu/
FlexX	BioSolveIT	http://www.biosolveit.de/FlexX/
Fred	OpenEyes	http://www.eyesopen.com/products/applications/fred.html
Glide	Schrödinger	http://www.schrodinger.com/Products/glide.html
Gold	CCDC	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
ICM	Molsoft	http://www.molsoft.com/products.html
LigandFit	Accelrys	http://www.accelrys.com/cerius2/c2ligandfit.html
Surflex	Biopharmics	http://www.biopharmics.com/products.html

[23]. Conversely, ligand flexibility is very well addressed within an upper limit of ~15 rotatable bonds. Three main options are possible to consider ligand flexibility in virtual screening: (1) an ensemble of previously-generated conformations is saved for each ligand that is docked as a rigid body to the binding site (e.g. Fred), (2) the ligand is built by an incremental assembly of fragments within the active site (e.g. FlexX), (3) the conformational space of the ligand is sampled by a Monte-Carlo (e.g. Glide) or a genetic-algorithm (e.g. Gold) procedure during the docking. Generally, several poses of the ligand are saved and scored by decreasing order of probability, assessed by a scoring function [24]. Scoring functions can be classified into three main categories [25]: empirical scoring functions, knowledge-based potentials and force-field methods (Table 12.3).

Table 12.3 Main scoring functions

Name	Category[a]	Ref.	Name	Category[a]	Ref.
AutoDock	A	55	Goldscore	A	76
Bleep	B	70	Hint	C	77
Chemscore	C	71	Ligscore	C	45
Cscore	C	72	Ludi	C	78
Dock	A	43	PLP	C	79
DrugScore	B	56	Pmf	B	61
FlexX	C	73	ScreenScore	C	80
Fresno	C	74	Smog	B	81
Glidescore	C	75	X-score	C	82

a) A = force-field; B = potential of mean force; C = empirical.

Empirical scoring functions (e.g. Ludi, Chemscore) use several terms describing properties known to be important in drug binding to construct a master equation for predicting binding affinity. Multilinear regression is used to optimize the coefficients to weight the computed terms using a training set of protein–ligand complexes for which both the binding affinity and an experimentally-determined 3-D structure is known. These terms generally describe polar and apolar interactions, loss of ligand flexibility (entropy) and eventually also desolvation effects. One major disadvantage of the empirical scoring functions is the need for a training set to derive the weight factors of the individual energy terms. It is therefore necessary to have available an empirical scoring function which performs well not only for proteins similar to that used in the training set. Potentials of mean force (e.g. PMF, Drugscore) simply encode structural information gathered from protein–ligand X-ray coordinates into Helmholtz free interaction energies of protein–ligand atom pairs. It is assumed that the more often a protein atom of type i and a ligand atom of type j are found in at certain distance rij, the more favorable is this interaction. The score is defined as the sum over all interatomic interactions of the protein–ligand complex. Advantages of this approach are that no fitting to ex-

perimentally-measured binding free energies of complexes in a training set is needed, and that solvation and entropic terms are treated implicitly. Force-field-based scoring functions (e.g. Dock energy score) are based on the non-bonded terms of a classical molecular mechanics force field (e.g. AMBER, CHARMM, etc.). A Lennard–Jones potential describes van-der-Waals interactions whereas the Coulomb energy describes the electrostatic components of the interactions. The main drawback of force-field calculations is the omission of the entropic component of the binding free energy. Therefore care should be taken not to overestimate the largest and most polar molecules that usually receive the highest enthalpy interaction scores.

The current accuracy of the most robust docking tools allow the prediction of the correct location of a ligand in ~75% of cases, provided that several solutions are saved [26]. A major problem is that the scoring function does not always (only in ~40–50% of the cases) predict the correct solution as the most probable one, which considerably complicates the analysis of docking results. There are numerous reasons for this limited accuracy [26], some of which are simple to correct (e.g. incorrect atom type for either the ligand or the protein), others pose more of a problem (e.g. accuracy of the protein 3-D structure, flexibility of the ligand, accuracy of the scoring function), and some are really tricky to overcome (protein flexibility, role of bound water). These limitations explain why it is very difficult to predict which docking tool and which scoring function is the most appropriate in the context of interest [27]. When docking is applied to a large database, the corresponding scoring function should be robust enough to rank putative hits by increasing binding free energy values [25]. This is still one of the biggest challenges in computational chemistry. Predicting binding free energy changes is possible in circumstances where a customized scoring function is applied to a series of congeneric ligands. However, for a database containing a large diversity of compounds, and for target families which have not been used for calibrating scoring functions, the accuracy obtained is usually limited [28]. From this observation, two sources of improvement are possible: (i) the design of more accurate scoring functions [29], and (ii) the design of smarter strategies to post-process docking outputs [27, 30]. The author personally favors the second option. The accuracy of scoring functions levelled off several years ago, for the simple reason that some unknown parameters (e.g. role of bound water, protein flexibility) remained difficult to predict whatever the physical principles used to derive a scoring function.

12.2.3
Data Post-processing

Acknowledging that scoring functions are far form being perfect, the best strategy to retrieve true positives from a virtual screen is first to detect false positives. Many strategies are possible. The simplest consists of re-scoring poses with additional scoring functions in the hope that consensus scoring [27, 30] will better identify true hits (top-ranked by several scoring functions) from decoys. Comparing hit rates between simple and consensus scoring should however be realized

on hit lists of comparative sizes [31]. However, customizing a consensus scoring scheme first requires knowledge of several and chemically-diverse true hits. Such data are not always available. Therefore, for poorly investigated targets, other strategies need to be designed. Topological filters can be used to filter out poses exhibiting steric or electrostatic mismatches between the ligand and its target [32]. Poses can also be minimized by third party software [33], hierarchical clustering [34], or analysis with Bayesian statistics[35]. In any case, the post-processing treatment should be simple enough to be reproducible for a wide array of targets. We have chosen to analyse docking outputs by looking at enrichment among true hits in pre-computed substructures/scaffolds [36]. This strategy presents the advantage of focusing more on scaffolds and how docking scores are distributed among them, and less on individual molecules. Last but not the least, selected hits should be browsed in 3-D target space for the ultimate selection: no algorithm yet outperforms the brain of an experienced modeller for such a task!

12.3
Retrospective Screening

12.3.1
Give it a Try

"Trying to dock a ligand into a GPCR-3D model has absolutely no relevance. At best it is only a silly idea". This definitive judgement from a referee commenting on our first paper on this topic in 2002 reflects the very strong reluctance of the scientific community to accept the simple idea that an homology model of a GPCR could be of any use in prioritizing ligand selection by structure-based screening. GPCR 3-D models were first used in a retrospective manner to discriminate known ligands from decoys embedded in a training dataset. The known existence of several conformations of a receptor (active(s) state(s), ground state) [37] further complicates the exploitation of GPCR structures because known ligands should be divided into different datasets according to their function. Starting from these very simple assumptions, we attempted to check whether rhodopsin-based homology models could discriminate known ligands from randomly-chosen drug-like molecules [38]. Three human receptors (dopamine D_3, vasopressin V_{1a}, muscarinic M_1 receptors) were chosen for that purpose for two main reasons: (i) numerous site-directed mutagenesis data could help to check whether the automated docking of well-known ligands was consistent with the experiment, and (ii) ligands for these three receptors were sufficiently diverse chemically to enable two datasets to be established: a training set (for customizing the receptor cavity) and a test set. It very soon appeared that threading a GPCR onto the 2.8-Å resolution X-ray structure of bovine rhodopsin [11] was not sufficient to produce a cavity in the transmembrane (TM) domain that would be large enough to accommodate any known ligand. First, a TM cavity customized to accommodate known ligands is a prerequisite. The approach we chose was first to overlay known li-

gands (either antagonists or agonists) according to their known pharmacophore (e.g. using FlexS [39]), dock them manually according to site-directed mutagenesis data, and then refine the receptor in the presence of all bound ligands. Antagonist-biased GPCR models could then discriminate known antagonists (belonging to chemical series different from those used to refine the receptor) from randomly-chosen decoys using multiple consensus scoring (Fig. 12.2a).
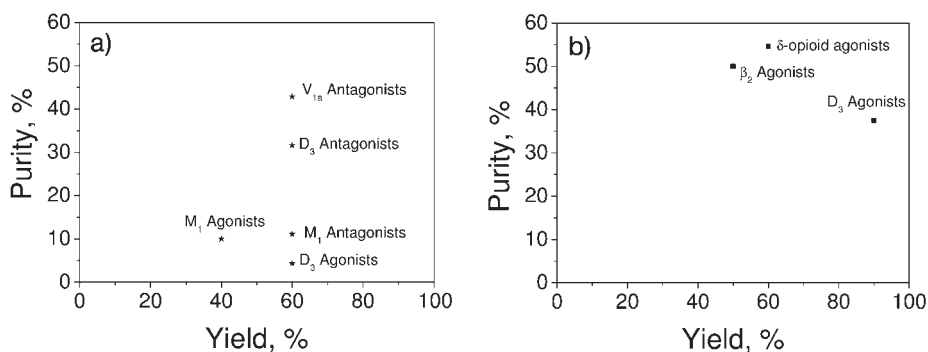
Fig. 12.2 Qualitative description of hit lists generated by virtual screening of a test dataset (10 true ligands, 990 randomly-chosen drug-like molecules) against homology models of three GPCRs. (a) Inactive state model, (b) active state model [38].

With the exception of the muscarinic M_1 receptor, the models generated of the dopamine D_3 and vasopressin $V1_a$ receptors were accurate enough to afford hit lists enriched in true antagonists (hit rate > 30%) and to recover a majority of known binders (yield > 60%) [38]. Such an approach was less successful in recovering known M_1 antagonists, probably because of the greater inaccuracy of the corresponding model. Moreover, the same strategy was totally unable to prioritize known agonists for selection (Fig. 12.2) implying that the X-ray structure of bovine rhodopsin in complex with the inverse agonist retinal was not a good template for modeling the active state of GPCRs. This thus led us to customize a different set of coordinates for active state models. Based on known data regarding the activation of aminergic GPCRs (for a review see [37]), we designed a computational protocol attempting to mirror what was known at that time: TM6 was rotated counter-clockwise from 30° (when viewed from the extracellular domain), a set of pre-aligned known agonists was manually docked according to experimental data, and the receptor was then refined in the presence of all bound ligands. This strategy allowed us to customize alternative 3-D active state models which were much more efficient in retrieving known full agonists (compare Fig. 12.2b with Fig. 12.2a) with hit rates and percentages of recovery of true binders very similar to those previously observed for antagonists.

12.3.2
Several Alternative Screening Strategies

Since our 2003 paper [38], several studies reporting the generation of GPCR models able to discriminate known ligands from randomly-chosen compounds have been published [40–42]. Interestingly, these studies achieve rather comparable hit rates and percentages of recovery of true binders despite very different strategies for generating 3-D receptor models.

Shacham et al. reported the PREDICT method [42] for modeling rhodopsin-independent GPCR 3-D structures and challenging them by virtual screening experiments. PREDICT is a multi-step computational protocol that identifies TM sequences, proposes alternative packing geometries for the seven helices (decoys) into 2-D space, optimizes the relative rotational orientation of the helices, converts the most likely decoys into a simplified 3-D representation, optimizes and clusters the best solutions, minimizes all-atoms models of each cluster representative and finally refines the most stable model by molecular dynamics. A virtual receptor–ligand model can be used to customize the binding site cavity bound to a known ligand. The method was applied to model antagonist-bound models of the dopamine D_2 receptor, the neurokinine NK_1 receptor and the neuropeptide Y (NPY) Y_1 receptor. DOCK4.0 [43] was used as a docking engine, in combination with CONFORT [44] as a conformation-generating tool, to dock test libraries in which a few known ligands (10–33) were seeded into a 10,000 drug-like compound library. Of all known binders 70–100% were recovered in the top 10% of compounds (ranked by the DOCK energy score) and 9–44-fold enrichments over random screening were reported. Interestingly, the same model of the dopamine D_2 receptor was able to retrieve both antagonists and full agonists although the latter were more difficult to distinguish from randomly-chosen drug-like compounds.

Gouldson et al. [41] reported a very different procedure in which both presumed active and inactive state models of the β_2-adrenergic receptor were obtained from experimental restraints (site-directed cross-kinking, engineered zinc binding, site-directed spin labeling, IR spectroscopy, cysteine accessibility) determined for class A GPCRs. Starting from the X-ray structure of bovine rhodopsin, the β_2 receptor was modeled and simulated in an explicit water environment. The use of two sets of experimental restraints (interatomic distance ranges) describing either an active or an inactive state model in MD simulation protocols afforded two sets of coordinates for the receptor. A library consisting of 172 GPCR ligands, in which β-adrenergic ligands (agonists and antagonists) were equally mixed with other GPCR ligands, was docked using LigandFit [45] and Ligscore as a scoring function. Interestingly, the active-state model was significantly more efficient in recruiting known agonists than the inactive state model for known antagonists. Extracting the top 10% of compounds scored, ~ 60% of the known agonists could be effectively recovered whereas only 11–42% of the known antagonists could be found. Hit rates (25–50%) were comparable to those previously described in two earlier studies [38, 42] although decoys in the dataset are likely to overlap the chemical space covered by true binders in the present case.

Moro et al. proposed another interesting approach [46] in which a congeneric series of 106 antagonists of the human adenosine A3 receptors (pyrazolo-triazolo-pyrimidines) had been first docked by MOE [47] to a rhodopsin-based homology model and then scored by the 3-D QSAR CoMFA method [48]. Steric and electrostatic contour plots derived from the CoMFA study were complementary to the predicted shape of the TM cavity. A cross-validated correlation of 0.84 was obtained from 106 compounds and six principal components. The standard error of prediction (0.84 pKi) was much lower than that derived by fitting experimental biological activities with docking scores. The CoMFA model was able to predict the affinity of 17 new antagonists from the same chemical series with a prediction error of 0.87 pKi. One reason for the success of the approach is that a set of very similar and quite rigid antagonists were used. Applying this strategy to a set of more chemically-diverse and flexible ligands has still to be undertaken in order to generalize this post-docking scoring approach.

12.4
Prospective Screening

12.4.1
Screening Rhodopsin-based Ligand-biased Homology Models

The very first predictive study was conducted by Varady et al. on the identification of dopamine D_3 receptor ligands [49]. A homology model of the receptor was first obtained from the X-ray structure of bovine rhodopsin, and simulated by molecular dynamics in a fully hydrated phospholipid bilayer with the CHARMM program [50]. A 2 ns-trajectory of the full system was further used to cluster receptor conformations into four families out of which one representative conformer was chosen for docking studies. A hybrid method was applied to screen the NCI 3D database [51] comprising over 250,000 compounds. First a pharmacophore was derived out of a set of 10 known D_3 ligands (antagonists and partial agonists) and used as a query for retrieving a first hit list of 6727 ligands satisfying the pharmacophore constraints. This hit list was then docked to each of the above-described four conformations of the D_3 receptor using LigandFit [45] with the Cerius^2 dockscore. A total of 2478 compounds ranked within the top 30% of the library for at least two receptor conformations was selected. This second hit list was analyzed for its chemical diversity to select 1314 molecules showing a Tanimoto similarity index lower than 80% to any of the 10 known D_3 receptor ligands. Twenty molecules were finally selected for biological evaluation; eight compounds exhibited submicromolar binding affinities out of which compounds 1–4 (Fig. 12.3) presented K_i values better than 100 nM.

Unfortunately, no functional assay was carried out to determine which type of ligand (antagonist, inverse agonist, partial agonist, and full agonist) was discovered by this elegant screening approach.

A similar although simpler strategy was used by Evers et al. [52, 53] to identify a new neurokinin-1 receptor antagonist. Ligand-biased modeling of the receptor
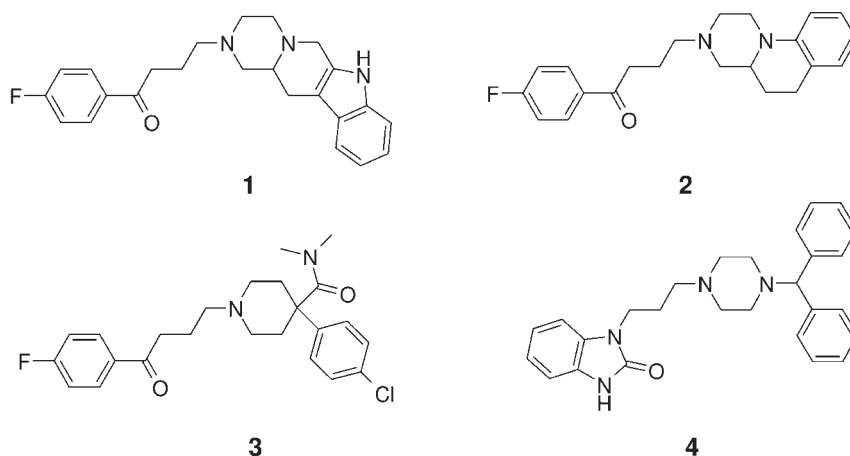
Fig. 12.3 Dopamine D$_3$ receptor antagonists identified by virtual screening [49].

was first realized with the MOBILE program. A set of crude homology models was first obtained from the crystal structure of bovine rhodopsin using MODEL-LER [54]. A known NK-1 antagonist (CP-96346) was then docked as a rigid body into each receptor model with AutoDock3.0 [55] Four poses satisfying the known interaction model were selected for generating a new set of receptor models scored according to DrugScore [56] to satisfy the ligand constraints. Selection and energy refinement of the best model according to DrugScore was finally achieved using the MAB force-field [57]. Over 800,000 commercially-available molecules were then screened in a multi-step protocol. A 1-D filtering discarding 50% of the starting compounds was first used to remove large and flexible structures. A topological filter was then assigned to retrieve 131,967 compounds satisfying a simple 2-D pharmacophore query. A 3-D search was subsequently applied to select 36,704 molecules satisfying a previously-derived 3-D pharmacophore hypothesis. Using an implicit definition of forbidden volumes, the size of the hit list was reduced to 11,109 compounds. Only these molecules were docked to the refined receptor model using FlexX-Pharm [58] and the DrugScore scoring function. A further restraint was used in the docking phase by imposing a specific H-bond to an important residue (Gln4.60) of the NK$_1$ receptor. Energy refinement of the 1000 best-ranked ligands and detailed visual inspection afforded the final selection of seven candidates. One of these seven molecules (compound **5**, Fig. 12.4) proved to bind to the NK1 receptor with an inhibition constant of 0.25 μM.

The same approach was also recently used to identify α_{1a} receptor antagonists from the Aventis compound collection [59]. Homology models of the α_{1a} receptor were derived and the one that best suited the known binding mode of a reference antagonist was selected for further database screening. A series of filters of increasing complexity were then applied to downsize the number of potential compounds to be docked. Topological and CATALYST [60] pharmacophore restraints were first used to select 22,950 compounds. A docking strategy (GOLD docking

5
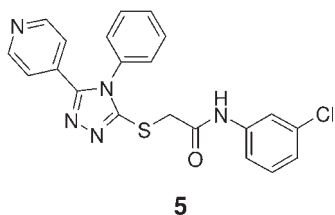
Fig. 12.4 Neurokinin NK$_1$ receptor antagonists identified by virtual screening [53].

and PMF [61] scoring) fine-tuned to recover 50 known α_{1a} antagonists seeded in 900 MDDR compounds, was applied to dock the first hit-list. The top-scored 300 compounds were then clustered according to UNITY [44] fingerprint similarity and a diverse subset of 80 molecules finally evaluated for receptor binding. Thirty-seven compounds (46%) exhibited K_i values lower than 10 µM, of which 24 molecules (30%) bound in the submicromolar range, 10 (12.5%) below 100 nM, and three (compounds **6–8**, Fig. 12.5) below 10 nM.

This remarkable hit rate is attributable both to the quality of the homology model for the target GPCR and the likely enrichment in ligands for aminergic receptors in the Aventis collection. However, smart screening is still necessary to distinguish true binders from promiscuous GPCR ligands although the selectivity of identified α_{1a} ligands was not reported in the report. The proposed procedure, like that reported by Varady [49] and Evers [53] presents the advantage of customizing the receptor model that better suits the restraints imposed by a ligand of known binding mode. Pharmacophore-based filtering followed by 3-D docking affords high-value potential hits. A major drawback of this strategy is that preliminary knowledge about known ligands is necessary to fine-tune the receptor model. Moreover, the choice of a relevant pharmacophore hypothesis must be undertaken

6

7

8
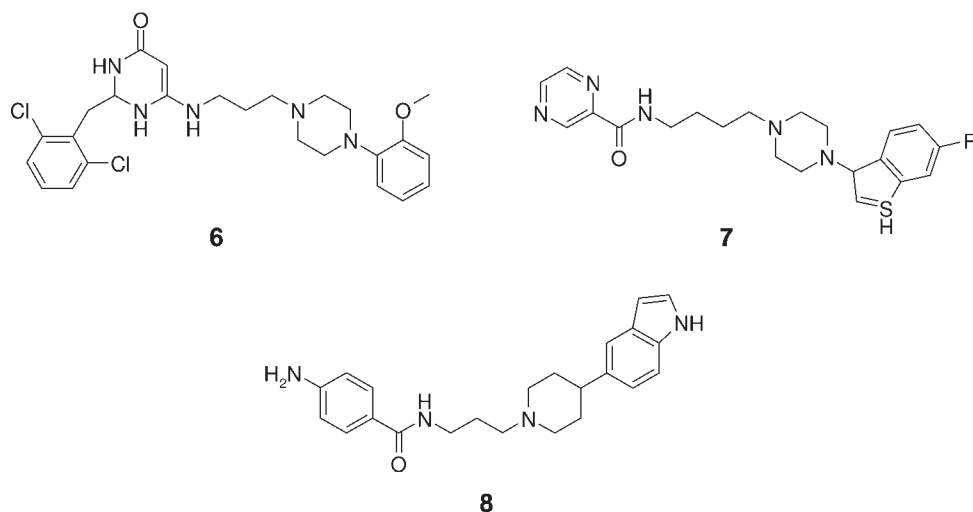
Fig. 12.5 Adrenergic $_{1a}$ receptor antagonists identified by virtual screening [59].

without too much bias in the selection of hits too close to the ligands used to customize the pharmacophore hypothesis and the receptor model. A good balance between focus and innovation is thus the key determinant of this approach.

12.4.2
Screening *Ab Initio* Models

Becker et al. [62] recently described the use of rhodopsin-independent models generated by the PREDICT method [42]. Starting with 1.6 million commercially-available compounds, DOCK4.0 was used to dock ~ 10 % of the full dataset to five class A GPCRs (5-HT$_{1A}$, NK$_1$, 5-HT$_4$, D$_2$, CCR$_3$). The top 10 %-ranked compounds passing an in-house topology filter are then re-scored using several scoring functions (Dock scores, Cscore, CHARMM). A series of unpublished cut-off values for each score was used to reduce the size of the hit list. The remaining compounds were then filtered by using a 3-D principle component analysis based on the 3-D properties of docked solutions. Molecules describing the same 3-D space as known binders were finally retrieved and clustered by diversity to afford a list of ~ 100 representative virtual hits for each receptor. Impressively, hit rates between 12 and 21 % were constantly achieved at a 5 µM cut-off value for four out of five targets. Although neither the models nor the structures of the hits identified were published, the authors reported the identification of new structures (antagonists and partial agonists) with low nanomolar binding affinities.

12.4.3
A Few Difficult Screening Scenarios

We recently challenged both GPCR 3-D models and receptor-based screening methods in two very difficult projects. The first was aimed at identifying, by serial screening, compounds exhibiting promiscuous binding to two dopamine receptor subtypes (D$_1$, D$_2$). To complicate the screening procedure even more, the search was directed towards compounds showing opposite functional effects on both receptors, full agonists at the D$_1$ subtype and antagonists at the D$_2$ subtype. We therefore generated two homology models of both receptors as previously described [38]. The dopamine D$_1$ receptor was first modeled in its presumed "active state" model [38] and customized to discriminate a set of 10 known dopamine D$_1$ agonists from randomly-chosen drug-like compounds. The dopamine D$_2$ receptor was modeled in its presumed "inactive state" model and refined according to our previously-described protocol, in presence of several known D$_2$ antagonists. This receptor model was also able to discriminate known D$_2$ antagonists from decoys. A library of over 2.1 million molecules from 20 different suppliers was then screened according to a sequential protocol comprised of filters of increasing complexity applied to decreasing numbers of compounds. OpenEye Filter program [63] was first used to select 310,000 drug-like compounds. A total of 40,000 molecules was then selected from this stock by applying a simple 2-D pharmacophore query. The hit list was reduced to 1834 compounds satisfying a 3-D UNITY [64]

query constructed from a set of 10 known dopamine D_1 agonists. All these molecules were docked with GOLD to the active state model of the D_1 receptor and 533 compounds were retrieved by a consensus scoring scheme previously optimized to distinguish known D_1 agonists from randomly-chosen molecules. These 533 compounds were then docked with GOLD into the inactive state model of the dopamine D_2 receptor using again a previously customized consensus scoring strategy known to discriminate known D_2 antagonists from decoys. A total of 200 putative hits was finally classified by ClassPharmer [83] to prioritize 17 scaffolds and at last one representative for each family. These 17 molecules were purchased and tested for binding to both D_1 and D_2 receptor subtypes.
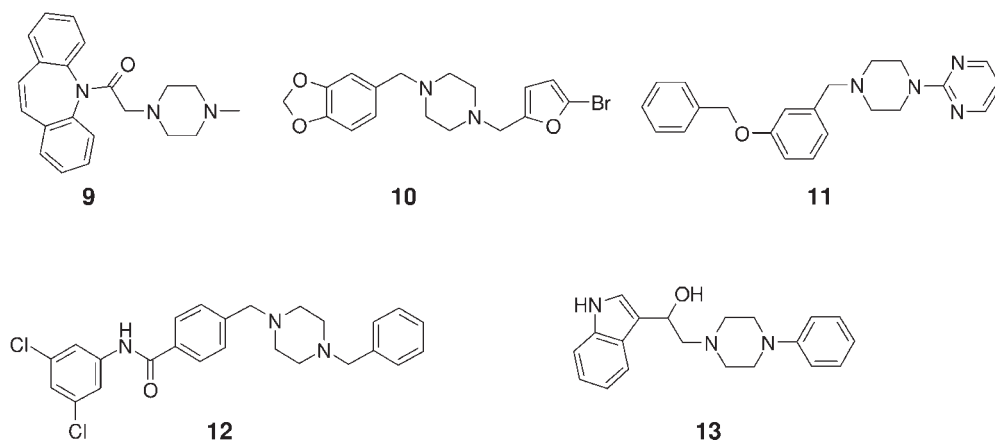
Fig. 12.6 Dopamine D_1/D_2 receptor ligands identified by serial screening.

Five of these 17 compounds bound to one of the two receptors with a K_i lower than 50 μM. All five compounds (**9–13**, Fig. 12.6) bound to the D_2 receptor whereas only three (**10**, **12**, **13**) also bound to the D_1 receptor. Strikingly, none of the five validated hits presented the expected functional profile (D_1 agonist and D_2 antagonist) when tested in a calcium assay and a cAMP gene reporter assay at both receptors recombinantly expressed in HEK293 cells (unpublished data). Although the screening was realized first on a putative "active state" model of the D_1 receptor and then on a "inactive state" model of the D_2 receptor, all D_1 ligands were neutral antagonists and three out of the five D_2 ligands (**11**, **12**, **13**) were full agonists. As a consequence, the panel of validated hits exhibited different profiles but not the required one. By looking at D_1 agonists/D_2 antagonists, we finally identified two micromolar D_1 antagonists/D_2 agonists (compounds **12**, **13**).

Another difficult issue has been to find non-peptide agonists for GPCRs whose endogenous ligands are natural peptides [65]. As part of a project aimed at identifying new CCR5 receptor antagonists as putative antiviral agents [66], we screened a composite library of ~ 45,000 drug-like compounds against a rhodopsin-based model of the CCR5 receptor and found new CCR5 non-peptide agonists! (Fig. 12.7).
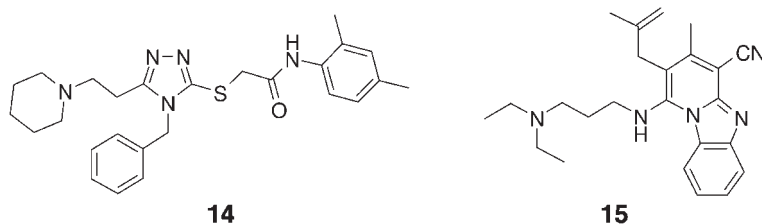
Fig. 12.7 Chemokine CCR5 receptor agonists identified by virtual screening.

It should be stated that the receptor model differed from the X-ray structure of bovine rhodopsin at TM2 in which a specific TXP motif (X being any amino acid) present in most chemokine receptors induced a specific bend of the helix which is not present in rhodopsin [67]. Although the model was refined in the presence of several CCR5 antagonists embedded in the receptor model according to experimental data [68] as previously described [38], docking in parallel with GOLD and SURFLEX and post-processing the common hits to the top 15% ranked compounds with ClassPharmer, afforded a hit list of 63 compounds of which 11 molecules exhibited detectable binding to the CCR5 receptor. Surprisingly, all of them were shown to be agonists, the most two potent hits (compounds **14**, **15**; Fig. 12.7) exhibited EC_{50} values of 22 and 17 μM, respectively, in a aequorin-based functional assay.

In both cases, it was surprising to find GPCR agonists by screening a rhodopsin-based model as we [38] and others [69] previously stated that the bovine rhodopsin model is only likely to be a good template for discovering inverse agonists and neutral antagonists. The current example is in agreement with Gouldson's work on the β_2 adrenoreceptor suggesting that agonists may bind to both active and inactive state models [41]. It demonstrates that identifying full agonists from inactive-state models is nevertheless feasible, although more potent agonists and higher hit rates may have been achieved by screening the appropriate active state model. Moreover, it opens the door to the virtual screening of "difficult targets" (e. g. class B GPCRs) with the objective of identifying non-peptide agonists.

12.5
Conclusions

Structure-based virtual screening is an emerging computational method showing some promise when applied to high-resolution X-ray structures. However, applying this technique to GPCRs first requires that several conditions are satisfied. The most important is to refine a receptor model which takes into account as much ligand information as possible. All successful approaches in the field of GPCR ligand design have relied on a receptor model that has been strongly biased by the structure of known ligands. It is very unlikely that 3-D models automatically derived from the X-ray structure of bovine rhodopsin will be of helpful. Sec-

ond, careful selection of the most appropriate ligand library is essential. There is always a temptation to screen large libraries in order to find totally new and unexpected chemotypes. Our experience suggests the opposite approach. The most difficult aspect of library design/selection is to reach a optimum balance between focus and innovation. Smart GPCR-targeted libraries and/or privileged structures will need to follow this line to really enrich virtual hit lists in true actives. Last, methods of mining docking results still need to be considerably improved. Inasmuch as scoring functions are far from being perfect, more importance should be given to the distribution of docking scores in chemical space and less to individual values. Recent progress in HTS data analysis is likely to benefit the analysis of virtual screening data.

References

1 Klabunde, T. and G. Hessler. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **2002**; *3*: 928–944.

2 Clardy, J. and C. Walsh. Lessons from natural molecules. *Nature* **2004**; *432*: 829–837.

3 Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol Ther* **2004**; *103*: 21–80.

4 Wise, A., K. Gearing, and S. Rees. Target validation of G-protein coupled receptors. *Drug Discovery Today* **2002**; *7*: 235–246.

5 Kubinyi, H. Chance favors the prepared mind – from serendipity to rational drug design. *J Recept Signal Transduct Res* **1999**; *19*: 15–39.

6 Wermuth, C. G. Selective optimization of side activities: another way for drug discovery. *J Med Chem* **2004**; *47*: 1303–1314.

7 Vassilatis, D. K., J. G. Hohmann, H. Zeng, F. Li, J. E. Ranchalis, M. T. Mortrud, A. Brown, S. S. Rodriguez, J. R. Weller, A. C. Wright, J. E. Bergmann, and G. A. Gaitanaris. The G protein-coupled receptor repertoires of human and mouse. *Proc Natl Acad Sci USA* **2003**; *100*: 4903–4908.

8 Lundstrom, K. Structural genomics on membrane proteins: mini review. *Comb Chem High Throughput Screen* **2004**; *7*: 431–439.

9 Fredriksson, R., M. C. Lagerstrom, L. G. Lundin, and H. B. Schioth. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **2003**; *63*: 1256–1272.

10 Gasteiger, J. and T. Engel. *Chemoinfomatics. A Textbook*. Wiley-VCH, Weinheim, **2003**.

11 Palczewski, K., T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. L. Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **2000**; *289*: 739–745.

12 Anderson, A. C. The process of structure-based drug design. *Chem Biol* **2003**; *10*: 787–797.

13 Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**; *432*: 862–865.

14 Baurin, N., R. Baker, C. Richardson, I. Chen, N. Foloppe, A. Potter, A. Jordan, S. Roughley, M. Parratt, P. Greaney, D. Morley, and R. E. Hubbard. Drug-like annotation and duplicate analysis of a 23-supplier chemical database totalling 2.7 million

compounds. *J Chem Inf Comput Sci* **2004**; *44*: 643–651.

15 GASTEIGER, J. and T. ENGEL. *Chemoinfomatics. A Textbook*. Wiley-VCH, Weinheim, **2003**.

16 CHARIFSON, P. S. and W. P. WALTERS. Filtering databases and chemical libraries. *J Comput Aided Mol Des* **2002**, *16*, 311–323.

17 HANN, M., B. HUDSON, X. LEWELL, R. LIFELY, L. MILLER, and N. RAMSDEN. Strategic pooling of compounds for high-throughput screening. *J Chem Inf Comput Sci* **1999**; *39*: 897–902.

18 MUEGGE, I. Selection criteria for drug-like compounds. *Med Res Rev* **2003**; *23*: 302–321.

19 LIPINSKI, C. A., F. LOMBARDO, B. W. DOMINY, and P. J. FENEY. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Drug Delivery Rev* **1997**; *23*: 3–25.

20 SEIDLER, J., S. L. MCGOVERN, T. N. DOMAN, and B. K. SHOICHET. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J Med Chem* **2003**; *46*: 4477–4486.

21 BENDER, A. and R. C. GLEN. Molecular similarity: a key technique in molecular informatics. *Org Biomol Chem* **2004**; *2*: 3204–3218.

22 BROOIJMANS, N. and I. D. KUNTZ. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **2003**; *32*: 335–373.

23 FERRARI, A. M., B. Q. WEI, L. COSTANTINO, and B. K. SHOICHET. Soft docking and multiple receptor conformations in virtual screening. *J Med Chem* **2004**; *47*: 5076–5084.

24 GOHLKE, H. and G. KLEBE. Statistical potentials and scoring functions applied to protein–ligand binding. *Curr Opin Struct Biol* **2001**; *11*: 231–235.

25 KITCHEN, D. B., H. DECORNEZ, J. R. FURR, and J. BAJORATH. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* **2004**; *3*: 935–949.

26 KELLENBERGER, E., J. RODRIGO, P. MULLER, and D. ROGNAN. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**; *57*: 225–242.

27 BISSANTZ, C., G. FOLKERS, and D. ROGNAN. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J Med Chem* **2000**; *43*: 4759–4767.

28 FERRARA, P., H. GOHLKE, D. J. PRICE, G. KLEBE, and C. L. BROOKS, 3rd. Assessing scoring functions for protein–ligand interactions. *J Med Chem* **2004**; *47*: 3032–3047.

29 WANG, R., Y. LU, X. FANG, and S. WANG. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein–ligand complexes. *J Chem Inf Comput Sci* **2004**; *44*: 2114–2125.

30 CHARIFSON, P. S., J. J. CORKERY, M. A. MURCKO, and W. P. WALTERS. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* **1999**; *42*: 5100–5109.

31 XING, L., E. HODGKIN, Q. LIU, and D. SEDLOCK. Evaluation and application of multiple scoring functions for a virtual screening experiment. *J Comput Aided Mol Des* **2004**; *18*: 333–344.

32 STAHL, M. and H. J. BOHM. Development of filter functions for protein–ligand docking. *J Mol Graph Model* **1998**; *16*: 121–132.

33 TAYLOR, R. D., P. J. JEWSBURY, and J. W. ESSEX. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J Comput Chem* **2003**; *24*: 1637–1656.

34 PAUL, N. and D. ROGNAN. ConsDock: A new program for the consensus analysis of protein–ligand interactions. *Proteins* **2002**; *47*: 521–533.

35 KLON, A. E., M. GLICK, and J. W. DAVIES. Combination of a naive Bayes classifier with consensus scoring improves enrichment of high-throughput docking results. *J Med Chem* **2004**; *47*: 4356–4359.

36 NICOLAOU, C. A., S. Y. TAMURA, B. P. KELLEY, S. I. BASSETT, and R. F. NUTT. Analysis of large screening

data sets via adaptively grown phyloge-
netic-like trees. *J Chem Inf Comput Sci*
2002; *42*: 1069–1079.

37 Gether, U. Uncovering molecular
mechanisms involved in activation of
G protein-coupled receptors. *Endocrinol
Rev* **2000**; *21*: 90–113.

38 Bissantz, C., P. Bernard, M. Hibert,
and D. Rognan. Protein-based virtual
screening of chemical databases. II. Are
homology models of G-Protein Coupled
Receptors suitable targets? *Proteins*
2003; *50*: 5–25.

39 Lemmen, C., T. Lengauer, and G. Klebe.
FLEXS: a method for fast flexible ligand
superposition. *J Med Chem* **1998**; *41*:
4502–4520.

40 Ashton, M., M. H. Charlton,
M. K. Schwarz, R. J. Thomas, and
M. Whittaker. The selection and design
of GPCR ligands: from concept to the
clinic. *Comb Chem High Throughput
Screen* **2004**; *7*: 441–452.

41 Gouldson, P. R., N. J. Kidley,
R. P. Bywater, G. Psaroudakis,
H. D. Brooks, C. Diaz, D. Shire, and
C. A. Reynolds. Toward the active con-
formations of rhodopsin and the beta2-
adrenergic receptor. *Proteins* **2004**; *56*:
67–84.

42 Shacham, S., Y. Marantz, S. Bar-Haim,
O. Kalid, D. Warshaviak, N. Avisar,
B. Inbal, A. Heifetz, M. Fichman,
M. Topf, Z. Naor, S. Noiman, and
O. M. Becker. PREDICT modeling and
in-silico screening for G-protein coupled
receptors. *Proteins* **2004**; *57*: 51–86.

43 Ewing, T. J., S. Makino, A. G. Skill-
man, and I. D. Kuntz. DOCK 4.0: search
strategies for automated molecular
docking of flexible molecule databases.
J Comput Aided Mol Des **2001**; *15*: 411–
428.

44 SYBYL; 6.91 ed.; TRIPOS, Assoc., Inc.,
St Louis, MO, USA, **2003**.

45 Venkatachalam, C. M., X. Jiang,
T. Oldfield, and M. Waldman. Ligand-
Fit: a novel method for the shape-direc-
ted rapid docking of ligands to protein
active sites. *J Mol Graph Model* **2003**; *21*:
289–307.

46 Moro, S., P. Braiuca, F. Deflorian,
C. Ferrari, G. Pastorin, B. Cacciari,

P. G. Baraldi, K. Varani, P. A. Borea,
and G. Spalluto. Combined target-
based and ligand-based drug design
approach as a tool to define a novel 3D-
pharmacophore model of human A3
adenosine receptor antagonists: pyra-
zolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine
derivatives as a key study. *J Med Chem*
2005; *48*: 152–162.

47 Chemical Computing group: Montreal,
Quebec, Canada.

48 Cramer, I., R.D., D. E. Pattersson, and
J. D. Bunce. Comparative molecular
fields analysis (CoMFA). 1. Effect of
shape on binding of steroids to carrier
proteins. *J Am Chem Soc* **1988**; *110*:
5959–5967.

49 Varady, J., X. Wu, X. Fang, J. Min,
Z. Hu, B. Levant, and S. Wang. Mole-
cular modeling of the three-dimensional
structure of dopamine 3 (D3) subtype
receptor: discovery of novel and potent
D3 ligands through a hybrid pharmaco-
phore- and structure-based database
searching approach. *J Med Chem* **2003**;
46: 4377–4392.

50 Brooks, B., R. Bruccoleri, B. D. Olaf-
son, D. States, S. Swaminathan, and
M. Karplus. CHARMM: A program for
macromolecular energy, minimization,
and dynamics calculations. *J Comput
Chem* **1983**; *4*: 187–217.

51 Milne, G. W., M. C. Nicklaus,
J. S. Driscoll, S. Wang, and
D. Zaharevitz. National Cancer
Institute Drug Information System 3D
database. *J Chem Inf Comput Sci* **1994**;
34: 1219–1224.

52 Evers, A. and G. Klebe. Ligand-sup-
ported homology modeling of g-protein-
coupled receptor sites: models sufficient
for successful virtual screening. *Angew
Chem Int Ed Engl* **2004**; *43*: 248–251.

53 Evers, A. and G. Klebe. Successful vir-
tual screening for a submicromolar an-
tagonist of the neurokinin-1 receptor
based on a ligand-supported homology
model. *J Med Chem* **2004**; *47*: 5381–
5392.

54 Sali, A. and T. L. Blundell. Compara-
tive protein modelling by satisfaction of
spatial restraints. *J Mol Biol* **1993**; *234*:
779–815.

55 Morris, G. M., D. S. Goodsell, R. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function. *J Comp Chem* **1998**; *19*: 1639–1662.

56 Gohlke, H., M. Hendlich, and G. Klebe. Knowledge-based scoring function to predict protein–ligand interactions. *J Mol Biol* **2000**; *295*: 337–356.

57 Gerber, P. R. and K. Muller. MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J Comput Aided Mol Des* **1995**; *9*: 251–268.

58 Hindle, S. A., M. Rarey, C. Buning, and T. Lengaue. Flexible docking under pharmacophore type constraints. *J Comput Aided Mol Des* **2002**; *16*: 129–149.

59 Evers, A. and T. Klabunde. Structure-based drug discovery using GPCR homolgy modelling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J Med Chem* **2005** (in press).

60 Accelrys Inc., San Diego, CA, USA, **2003**.

61 Muegge, I. and Y. C. Martin. A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *J Med Chem* **1999**; *42*: 791–804.

62 Becker, O. M., Y. Marantz, S. Shacham, B. Inbal, A. Heifetz, O. Kalid, S. Bar-Haim, D. Warshaviak, M. Fichman, and S. Noiman. G protein-coupled receptors: in silico drug discovery in 3D. *Proc Natl Acad Sci USA* **2004**; *101*: 11304–11309.

63 OpenEye Scientific software, Santa Fe, NM 87507.

64 TRIPOS, Assoc., Inc., St Louis, MO, USA.

65 Jones, R. M., P. D. Boatman, G. Semple, Y. J. Shin, and S. Y. Tamura. Clinically validated peptides as templates for de novo peptidomimetic drug design at G-protein-coupled receptors. *Curr Opin Pharmacol* **2003**; *3*: 530–543.

66 Maeda, K., H. Nakata, H. Ogata, Y. Koh, T. Miyakawa, and H. Mitsuya. The current status of, and challenges in, the development of CCR5 inhibitors as therapeutics for HIV-1 infection. *Curr Opin Pharmacol* **2004**; *4*: 447–452.

67 Govaerts, C., C. Blanpain, X. Deupi, S. Ballet, J. A. Ballesteros, S. J. Wodak, G. Vassart, L. Pardo, and M. Parmentier. The TXP motif in the second transmembrane helix of CCR5. A structural determinant of chemokine-induced activation. *J Biol Chem* **2001**; *276*: 13217–13225.

68 Dragic, T., A. Trkola, D. A. Thompson, E. G. Cormier, F. A. Kajumo, E. Maxwell, S. W. Lin, W. Ying, S. O. Smith, T. P. Sakmar, and J. P. Moore. A binding pocket for a small molecule inhibitor of HIV-1 entry within the transmembrane helices of CCR5. *Proc Natl Acad Sci USA* **2000**; *97*: 5639–5644.

69 Archer, E., B. Maigret, C. Escrieut, L. Pradayrol, and D. Fourmy. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol Sci* **2003**; *24*: 36–40.

70 Mitchell, J. B., R. A. Laskowski, A. Alex, and J. M. Thornton. BLEEP: potential of mean force describing protein–ligand interactions. II. Calculation of binding energies and comparison with experimental data. *J Comp Chem* **1999**; *20*: 117–1185.

71 Eldridge, M., C. W. Murray, T. A. Auton, G. V. Paolini, and R. P. Lee. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput-Aided Mol Des* **1997**; *11*: 425–445.

72 Clark, R. D., A. Strizhev, J. M. Leonard, J. F. Blake, and J. B. Matthew. Consensus scoring for ligand/protein interactions. *J Mol Graph Model* **2002**; *20*: 281–295.

73 Rarey, M., B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* **1996**; *261*: 470–489.

74 Rognan, D., S. L. Laumoeller, A. Holm, S. Buus, and V. Tschinke. Predicting binding affinities of protein ligands from three-dimensional coordinates: application to peptide binding to

class I major histocompatibility proteins. *J Med Chem* **1999**; *42*: 4650–4658.

75 Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**; *47*: 1739–1749.

76 Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. Improved protein–ligand docking using GOLD. *Proteins* **2003**; *52*: 609–623.

77 Cozzini, P., M. Fornabaio, A. Marabotti, D. J. Abraham, G. E. Kellogg, and A. Mozzarelli. Simple, intuitive calculations of free energy of binding for protein–ligand complexes. 1. Models without explicit constrained water. *J Med Chem* **2002**; *45*: 2469–2483.

78 Böhm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J Comput-Aided Mol Des* **1994**; *8*: 243–256.

79 Gehlhaar, D. K., G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, and S. T. Freer. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* **1995**; *2*: 317–324.

80 Stahl, M. and M. Rarey. Detailed analysis of scoring functions for virtual screening. *J Med Chem* **2001**; *44*: 1035–1042.

81 Ishchenko, A. V. and E. I. Shakhnovich. SMall Molecule Growth 2001 (SMoG2001): an improved knowledge-based scoring function for protein–ligand interactions. *J Med Chem* **2002**; *45*: 2770–2780.

82 Wang, R., L. Lai, and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput-Aided Mol Des* **2002**; *16*: 11–26.

83 Bioreason Inc., Sante Fe, New Mexico, USA.

Subject Index