Sequence-Derived Three-Dimensional Pharmacophore Models for G-Protein-Coupled Receptors and Their Application in Virtual Screening

Thomas Klabunde,* Clemens Giegerich, and Andreas Evers

Research & Development, Drug Design, Sanofi-Aventis Deutschland GmbH, D-65926 Frankfurt am Main, Germany

Received February 3, 2009

G-protein-coupled receptors (GPCRs) comprise a large protein family of significant past and current interest of pharmaceutical research. X-ray crystallography and molecular modeling combined with site-directed mutagenesis studies suggest that most family A GPCRs share a small-molecule binding site located in the outer part of the seven-transmembrane (7TM) bundle. Here we describe an automated method to derive sequence-derived three-dimensional (3D) pharmacophore models capturing the key elements for addressing this binding site by a small-molecule ligand. We have generated structure-based pharmacophore models from 10 homology models and 3 X-ray structures of receptor−ligand complexes. These 13 pharmacophores have been dissected into 35 different single-feature pharmacophore elements, each associated with a sequence motif or *chemoprint*, describing its molecular interaction partner(s) in the receptor. Subsequently, the protein sequences of 270 GPCRs have been searched for the presence of chemoprints and the appropriate single-feature pharmacophores have been assembled into three- to seven-feature 3D-pharmacophore models for each human family A GPCR. These models can be applied for virtual screening and for the design of subfamily directed libraries. A case study demonstrates the successful application of this approach for the identification of potent agonists for the complement component 3a receptor 1 (C3AR1) by virtual screening.

Introduction

G-protein-coupled receptors (GPCRs[a]) form a large protein family that plays an important role in many physiological and pathophysiological processes. GPCRs are located on the cell surface and are responsible for the transduction of an endogeneous signal into an intracellular response.[1] The natural ligands of this receptor superfamily are extremely diverse, comprising peptide and protein hormones, biogenic amines, nucleosides and nucleotides, lipids and eicosanoids, and others. Ligand binding causes conformational changes of the receptor transferring the signal to intracellular trimeric guanine nucleotide binding regulatory proteins (G-proteins). These G-proteins act as intracellular messengers, leading to an increase (or descrease) of cAMP or calcium or triggering the phosphorylation and internalization of the receptor.

Historically, the discovery of drugs acting at GPCRs has been extremely successful. Looking at the U.S. market only, 13 of 37 drug molecules with sales of more than 1 billion U.S. dollars each in 2007 act through GPCRs.[2] In 2007 the total sales of these 13 GPCR-directed blockbusters sum up to 23.5 billion U.S. dollars for the U.S. only. Furthermore, there are still approximately 100 GPCRs representing a valid or promising target of current pharmaceutical research. For many of these receptors, most major pharmaceutical companies currently profile drug candidates in preclinical studies or clinical trials or have compounds still undergoing chemical lead optimization. In addition, several GPCRs are still lacking a small-molecule chemical lead structure or even the physiological ligand (so-called orphan GPCRs).[3]

Up to now experimental three-dimensional (3D) structural information is available for four GPCRs: bovine rhodopsin,[4] the β2-adrenergic receptor,[5] the β1-adrenergic receptor,[6] and the adenosine A2A receptor.[7] The 3D-structures reveal that the outer half of the seven-transmembrane (7TM) bundle forms a small-ligand binding site in its center. In the adenosine A2A receptor this site extends further to the extracellular region of the receptor, showing significant molecular interactions with the second extracellular (ECL2) loop. On the basis of structural data and on the basis of molecular modeling studies supported by site-directed mutagenesis data, it has been suggested that all (family A) GPCR receptors share a small-molecule binding pocket located at a similar spatial position as retinal in rodopsin (or as carazolol in the β2-adrenergic receptor, respectively). This site has been proposed to be the switch for modulating the functional activity of the receptor by harboring agonists, inverse agonists, or neutral antagonists.[8]

In order to identify a chemical starting point targeting a GPCR of interest, academic research groups and pharmaceutical companies perform virtual screening in addition to random high-throughput screening.[9−15] Besides structure-based and ligand-based virtual screening approaches, chemogenomics-based virtual screening represents a possible approach, especially when ligand information for the target is insufficient or completely missing.[16] Following the chemogenomics paradigm "similar receptors bind similar ligands" experimental testing of ligands of a similar receptor on the target of interest is expected to increase the chance of identifying bioactive molecules compared to screening a random selection.[16,17] In order to describe the similarity of proteins, a classically applied similarity metric is full-sequence homology. However, as only the similarity of the putative ligand binding site is important for ligand recognition, other concepts have been presented that focus the comparison

* To whom correspondence should be addressed. Phone: +49 69 305 14355. Fax: +49 69 331399. E-mail: Thomas.Klabunde@sanofi-aventis.com.

[a] Abbreviations: GPCR, G-protein-coupled receptor; 3D, three-dimensional; C3AR1, complement component 3a receptor 1; AT1, angiotensin II receptor type 1; MOBILE, modeling binding sites including ligand information explicitly; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HYD, hydrophobic; HYDArom, hydrophobic aromatic; NI, negatively ionizable; PI, positively ionizable; RA, ring aromatic; SAR, structure−activity relationship; 7TM, seven-transmembrane; U2, urotensin-II; WDI, World Drug Index; MCH, melanin-concentrating hormone; ECL2, second extracellular loop.

10.1021/jm9001346 CCC: $40.75
Published on Web 04/17/2009

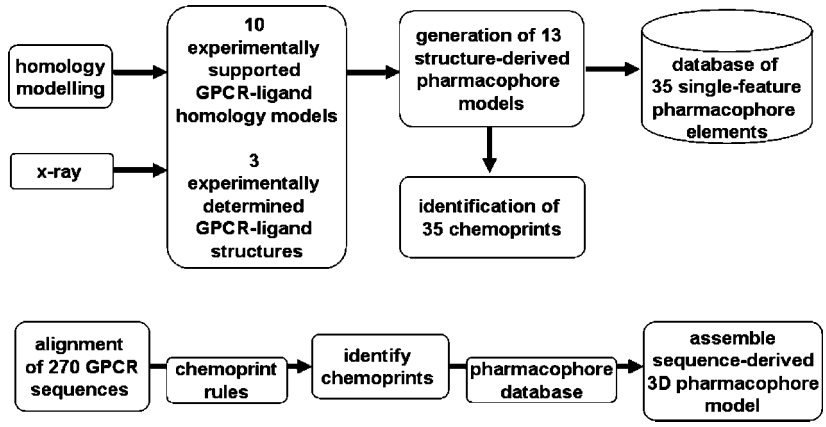


Figure 1. Flowchart describing the generation of sequence-derived 3D-pharmacophore models in a schematic manner. The top part shows the generation and analysis of structure-derived pharmacophore models for 13 reference GPCRs, resulting in 35 chemoprint rules and a database containing single-feature pharmacophore elements. The bottom part shows the generation of sequence-derived 3D-pharmacophore models for all family A GPCRs in an automated manner using the chemoprint rules and the pharmacophore element database derived from the 13 reference GPCRs. Details are given in the text.

Table 1. Structural Models of Receptor−Ligand Complexes Used To Generate Structure-Based Pharmacophore Models and To Derive Chemoprint Information[a]

GPCR	type	ligand	reference
β2 adrenergic receptor	X-ray structure	**1**, carazolol	5
β1 adrenergic receptor	X-ray structure	**2**, cyanopindolol	6
α1a adrenergic receptor	homology model	**3**, sertindol	27−29
5HT2A receptor	homology model	**4**, MDL100,907	in-house data
dopamine D2 receptor	homology model	**5**, HP873 metabolite	27, 30
muscarinic M1 receptor	homology model	**6**	31
angiotensin AT1 receptor	homology model	**7**, losartan	32
adenosine A2A receptor	X-ray structure	**8**, ZM241385	7
cannabinoid CB1 receptor	homology model	**9**, rimonabant	33
chemokine CCR2 receptor	homology model	**10**, RS-504393	34
chemokine CXCR2 receptor	homology model	**11**	35
EDG1 receptor	homology model	**12**, sphingosin-1-phosphate	36
NAR1 receptor	homology model	**13**, acifran	37

[a] Chemical structures of the ligands are shown in Scheme 1.

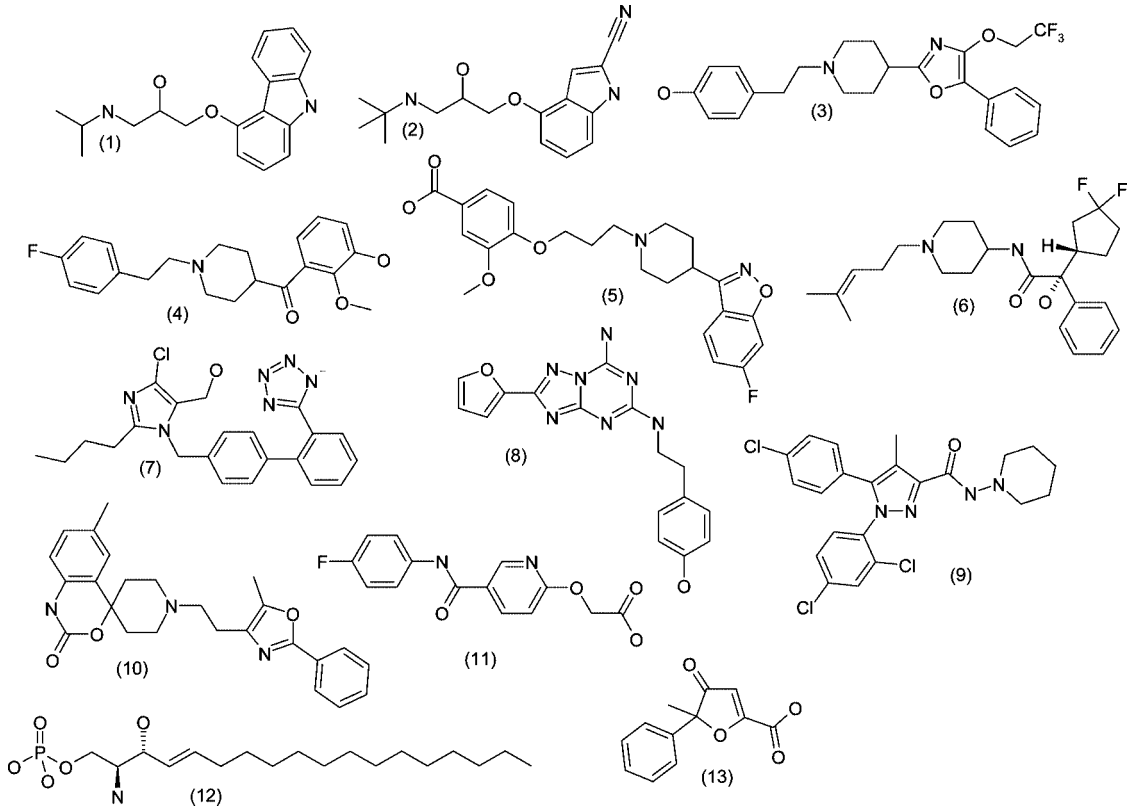
on the physicochemical properties of the amino acids facing the putative binding site.[18] Another view angle to define similarity is the comparison of common sequence motifs for recognition of ligand fragments, so-called *chemoprints* (e.g., Asp3.32 of biogenic amine receptors interacting with a positively ionizable nitrogen).[17] With 3D-structural information on GPCRs and their ligand complexes still being limited, homology modeling supported by site-directed mutagenesis data can be used to generate reliable 3D-structural information on GPCRs and their ligand complexes.[19] These experimentally supported homology models can be used as a source to derive chemoprints for GPCRs.

In this article we describe a novel automated approach for the generation of sequence-derived 3D-pharmacophore models for family A GPCRs (Figure 1). In the first step, we have generated 10 homology models for receptor−ligand complexes using the β2-adrenergic receptor as template. In these 10 homology models and in 3 receptor−ligand complex structures determined by X-ray crystallography the molecular interaction partners have been identified and 35 chemoprints have been captured. We hypothesize that GPCRs with the same (or similar) sequence motifs could recognize the same ligand feature or fragment at a similar spatial position. Guided by this hypothesis, we generated structure-based pharmacophore models from all 13 reference receptor structures. These reference pharmacophores have been dissected into 35 different single-feature pharmacophore elements. For each pharmacophore element the 3D-coordinates, the feature type, and the link to the corresponding amino acid sequence motif have been stored in a pharmacophore building block database. Subsequently, the protein sequences of all 270 family A GPCRs have been searched for the presence of chemoprints (Figure 1, lower part). The corresponding single-feature 3D-pharmacophore building blocks have been assembled into three- to seven-feature 3D-pharmacophore models for each GPCR in an automated manner. Many of these sequence-derived 3D-pharmacophore models have a degree of complexity allowing their application in virtual screening. Here we describe a successful application of a sequence-derived pharmacophore model generated using this Pharma^{3D} approach in virtual screening, leading to the discovery of potent agonists of the complement component 3a receptor 1 (C3AR1).

Methods and Results

Generation of Reference 3D Receptor−Ligand Complexes. We have generated 10 experimentally supported homology models of diverse GPCRs in complex with their small-molecule ligands (Table 1 and Scheme 1). For these 10 GPCRs sufficient site-directed mutagenesis data and ligand structure−activity data are available to guide the 3D-modeling procedure for the receptor−ligand complexes. We applied a modified version of the MOBILE approach[10,20] (**m**odeling **b**inding sites **i**ncluding **l**igand information **e**xplicitly), which models proteins by homology while explicitly including information about protein−ligand interactions as restraints. For each GPCR given in Table 1, molecular recognition data are translated into a topographical interaction model after analysis of all available mutagenesis data

Scheme 1



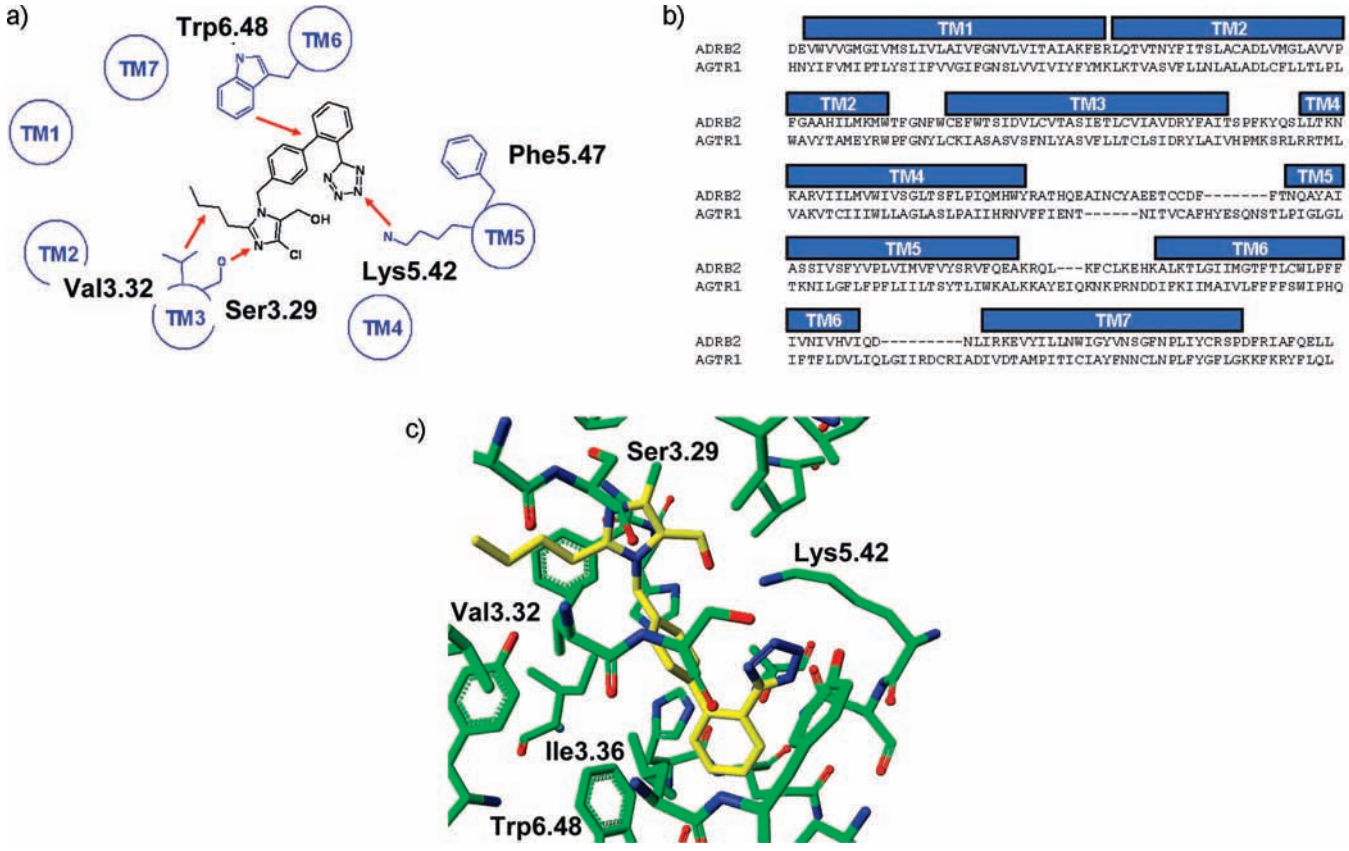


Figure 2. Generation of 3D-homology model for the AT1 receptor in complex with losartan using the MOBILE approach:[19] (a) topological interaction model for the molecular recognition of losartan in the AT1 receptor; (b) sequence alignment between the AT1 and the β2-adrenergic receptor used as template structure for homology model generation; (c) proposed 3D-model of the binding site of losartan in the AT1 receptor. Those residues interacting with the small-molecule ligand are labeled (nomenclature according to Weinstein−Ballesteros).

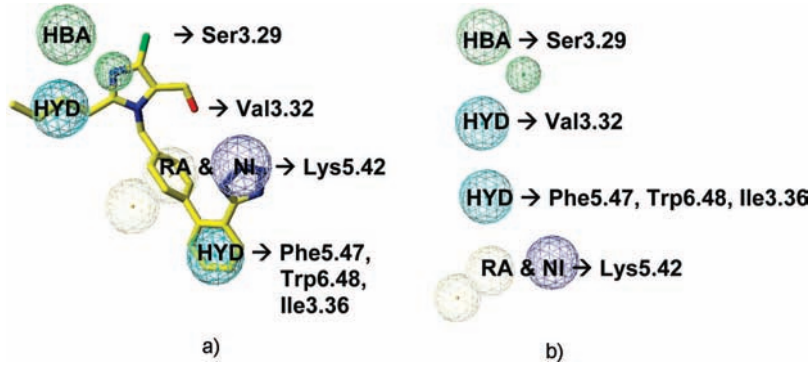


Figure 3. Generation of pharmacophore building block database, example AT1 receptor. Analysis of the complex of the AT1 receptor with losartan results in a structure-derived five-feature pharmacophore (a), which is in agreement with a ligand-derived pharmacophore based on losartan derivatives published by Langer et al.[26] The 3D-pharmacophore is dissected into single-feature pharmacophores, keeping their spatial position and associating each feature to the sequence-motif it is addressing (b).

and comparative affinity determinations based on ligand binding (as shown in Figure 2a for the angiotensin II receptor type 1 (AT1)). Following the MOBILE approach, 100 preliminary homology models of the target receptor are generated (with MOE)[21] using the X-ray structure of the β2-adrenergic receptor as structural template (Figure 2b). The target ligand is then docked into each of the 100 preliminary protein models (using GOLD),[22] imposing the known receptor−ligand interactions as docking constraints. The best-scored docking pose is considered for the next iteration. In this next iteration, 100 new homology models are generated, again using the X-ray structure of the β2-adrenergic receptor as template, now considering the docking pose from the previous step as additional restraint in the protein modeling procedure. The final model is composed by merging the amino acids with the best interactions to the reference ligand into one model. In order to relax the composed model, the entire binding pocket is minimized with the MMFF94 force field available in MOE, keeping ligand and protein residues flexible. Since knowledge about protein−ligand interactions is used to guide the GPCR modeling process, the resulting models are in agreement with these data, as shown in Figure 2c for the modeled complex of the AT1 receptor with bound losartan. The resulting receptor−ligand models are used for the identification of chemoprints (next section).

Identification of Chemoprints, Sequence Motifs for Ligand Recognition. The 3D-models of the generated receptor−ligand complexes have been analyzed to identify and capture the key receptor−ligand interaction pairs (e.g., Asp3.32 connected to a positively ionizable nitrogen in the ligand). In addition to the homology model-based complex structures, the experimental complex structures of the β2-adrenergic (PDB entry 2RH1), the β1-adrenergic (PDB entry 2VT4), and the adenosine A2A receptor (PDB entry 3EML) with bound antagonists have been analyzed.[5−7] Table 1 lists all reference GPCRs, together with the key references reporting the molecular recognition data (or describing the complexes determined by means of X-ray crystallography, respectively). Table 2 lists all receptor−ligand interaction pairs, which have been identified in the 13 reference complex models. We have coined the term *chemoprints* for these motifs in the GPCR sequence, which are hypothesized to be necessary to recognize and bind a specific ligand feature or fragment.[17] The numbering of the sequence motifs is given according to the nomenclature of Ballesteros−Weinstein, allowing identification of topological identical residues in GPCR sequences (the helix number is followed by an index representing the position of the residue relative to the most conserved residue in the helix, to which is arbitrarily assigned the index 50). It is evident that for some GPCRs the same molecular interaction pair is used (e.g., Asp3.32 in all biogenic amine GPCRs to recognize a positively ionizable nitrogen), resulting in some redundancy for several chemoprints. In total, 35 different chemoprints have been extracted from the reference models capturing the key interaction sites addressed by ligands within the 7TM binding site.

Generation of Structure-Based Reference Pharmacophore Models. Structure-based 3D-pharmacophore models have been generated for each of the 13 reference complexes using Catalyst. Toward this end, for each receptor−ligand complex the ligand has been extracted from the complex model and imported into Catalyst.[23] Using the pharmacophore feature dictionary of Catalyst (HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; HYD, hydrophobic; HYDArom, hydrophobic aromatic; NI, negatively ionizable; PI, positively ionizable; RA, ring aromatic), pharmacophore features are placed onto the ligand reflecting the receptor−ligand interaction pairs observed in the complex structures (Table 2).

Figure 3 illustrates the workflow using the AT1 receptor in complex with losartan as an example. The expected binding mode of losartan, which is supported by ligand structure−activity relationship (SAR) and site-directed mutagenesis data, is shown in Figure 2c. Analysis of the complex allows us to derive the structure-based pharmacophore model (Figure 3a). Each of the pharmacophore features is in agreement with ligand SAR data and is linked to an interaction partner in the GPCR binding site. Following the same workflow, structure-based pharmacophore models have been generated for all 13 GPCR−ligand complexes.

Single-Feature Pharmacophore Building Blocks. In order to generate a database of pharmacophore building blocks, each structure-based pharmacophore model has been dissected into single-feature pharmacophore elements (see AT1 example in Figure 3b). The dissection of each pharmacophore model in its elements has been performed using a Perl script: The script reads the pharmacophore hypothesis file generated by Catalyst and writes files containing the single-feature pharmacophore elements. In total, 35 different single-feature pharmacophores have been generated, each having the spatial position as found in 1 of the 13 receptor−ligand complexes. Each single-feature pharmacophore is associated with the chemoprint information, forming the respective interaction partner in the 7TM binding site.

Sequence-Derived Pharmacophore Models for Family A GPCRs. The single-feature pharmacophores with spatial positions as derived from the 13 reference complexes served

Table 2. Receptor−Ligand Interaction Pairs (Chemoprints) Derived from Complex Structures of Reference GPCRs (Residues from the ECL2 Loop Are Given in Parentheses)

chemoprint no.	reference GPCR	sequence motif	pharmacophoric feature
1	$\beta2/\beta1$ adrenergic	Asp3.32	pos ionizable
1	α1a adrenergic	Asp3.32	pos ionizable
1	5HT2A	Asp3.32	pos ionizable
1	D2	Asp3.32	pos ionizable
1	M1	Asp3.32	pos ionizable
2	$\beta2/\beta1$ adrenergic	Val3.33, Phe5.47, Phe6.52	hydrophobic aromatic
2	α1a adrenergic	Val3.33, Phe5.47, Tyr6.52	hydrophobic aromatic
2	5HT2A	Val3.33, Phe5.47, Phe6.52	hydrophobic aromatic
2	D2	Val3.33, Phe5.47, Phe6.52	hydrophobic aromatic
3	$\beta2/\beta1$ adrenergic	Ser5.42	hydrogen bond donor
4	$\beta2/\beta1$ adrenergic	Asn7.39	hydrogen bond acceptor
5	$\beta2$ adrenergic	Ala5.39, Phe6.51 (Phe5.32)	hydrophobic aromatic
5b	α1a adrenergic	Val5.39, Phe6.51, Met6.55	hydrophobic
6	α1a adrenergic	Phe2.64, Trp3.28	hydrophobic
7	5HT2A	Trp3.28	hydrophobic aromatic
8	5HT2A	Asn6.55	hydrogen bond acceptor
9	5HT2A	Ser5.43	hydrogen bond donor
10	D2	Leu2.64, Phe3.28	hydrophobic aromatic
11	D2	Ser7.39	hydrogen bond acceptor
12	M1	Ala3.36, Phe5.47, Trp6.48	hydrophobic aromatic
13	M1	Trp2.61, Trp3.28, Ala3.36	hydrophobic aromatic
14	M1	Asn6.52	hydrogen bond acceptor
15	M1	Ala5.43, Val6.55	hydrophobic
16	AT1	Lys5.42	neg ionizable, ring aromatic
17	AT1	Ile3.36, Phe5.47, Trp6.48	hydrophobic aromatic
18	AT1	Ser3.29	hydrogen bond acceptor
19	AT1	Val3.32	hydrophobic aliphatic
20	A2A	Asn6.55 (Glu5.30)	hydrogen bond donor
21	A2A	Asn6.55	hydrogen bond acceptor
22	A2A	Leu3.32, Trp6.48, Tyr6.51	hydrophobic aromatic
23	A2A	Ile7.39 (Phe5.29)	ring aromatic
24	CB1	Phe5.42	hydrophobic aromatic
25	CB1	Phe5.43, Trp6.48	hydrophobic aromatic
26	CB1	Phe2.61, Phe2.64	hydrophobic
27	CB1	Lys3.28	hydrogen bond acceptor
28	CCR2	Glu7.39	pos ionizable
29	CXCR2	Arg5.39	neg ionizable
30	CXCR2	Tyr6.51	hydrogen bond acceptor
31	CXCR2	Phe3.36, Tyr6.51	ring aromatic
32	CXCR2	Lys3.32, Glu7.39	hydrogen bond donor
33	EDG1	Arg3.28	neg ionizable
34	EDG1	Glu3.28	pos ionizable
35	NAR1	Arg6.55	neg ionizable

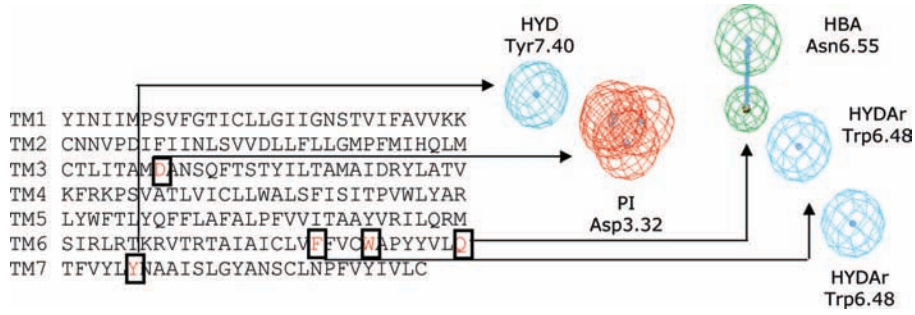


Figure 4. Sequence-derived 3D-pharmacophore model: example, MCH receptor. The sequence of the MCH receptor (7TM region) is searched for the presence of chemoprints. Several chemoprints can be identified, and for each chemoprint the respective 3D-pharmacophore feature is placed to assemble the sequence-derived 3D-pharmacophore. The redundancy of the positively ionizable (PI) feature with slightly varying spatial positions results from the fact that the feature is found in four different reference complexes and is thus stored in the 3D-pharmacophore database four times with slightly varying coordinates.

as building blocks for the generation of sequence-derived 3D-pharmacophores for all family A receptors. As mentioned above, we hypothesized that GPCRs with the same (or similar) sequence motif could recognize the same ligand feature or fragment at a similar spatial position. On the basis of a multiple alignment of the 7TM regions of all GPCRs, the sequences of all receptors are searched for the presence of chemoprints. Thus, a 35-digit chemoprint vector is generated for all 270 human family A GPCRs. A chemoprint "bit" is set on, not only if the exact chemoprint signature is present but even if similar amino acids are present, which would allow the same type of interaction as observed in the reference complex at a similar spatial position (for exact chemoprint conditions, see Table 3). If, for example, the interaction partner in the reference receptor is a valine residue forming a hydrophobic interaction to a ligand fragment (e.g., chemoprint 19), it is assumed a valine residue

Table 3. Chemoprint Conditions[a]

chemoprint no.	reference GPCR	sequence motif
1	$\beta2/\beta1$ adrenergic	D3.32
1	α1a adrenergic	D3.32
1	5HT2A	D3.32
1	D2	D3.32
1	M1	D3.32
2	$\beta2/\beta1$ adrenergic	(V3.33||A3.33||T3.33||I3.33||L3.33||G3.33||F3.33) & (F5.47||Y5.47||I5.47) & (F6.52||Y6.52||W6.52)
2	α1a adrenergic	(V3.33||A3.33||T3.33||I3.33||L3.33||G3.33) & (F5.47||Y5.47) & (F6.523||Y6.52)
2	5HT2A	(V3.33||A3.33||T3.33||I3.33||L3.33||G3.33) & (F5.47||Y5.47) & (F6.52||Y6.52)
2	D2	(V3.33||A3.33||T3.33||I3.33||L3.33||G3.33) & (F5.47||Y5.47) & (F6.52||Y6.52)
3	$\beta2/\beta1$ adrenergic	S5.42||T5.42
4	$\beta2/\beta1$ adrenergic	N7.39
5	$\beta2$ adrenergic	(F6.51||Y6.51) & (A5.39||G5.39||V5.39||I5.39||T5.39||L5.39||S5.39)
5b	α1a adrenergic	F6.51 & M6.55 & V5.39
6	α1a adrenergic	(W3.28||Y3.28||F3.28) & (F2.64||Y2.64) & !(F2.61||Y2.61||W2.61)
7	5HT2A	W3.28 & !(F2.64||Y2.64||W2.64)
8	5HT2A	N6.55||Q6.55
9	5HT2A	S5.43||T5.43
10	D2	(F3.28||Y3.28) & !(F2.61||Y2.61||W2.61) & (L2.64||I2.64||V2.64)
11	D2	T7.39||S7.39
12	M1	W6.48 & (F5.47||Y5.47) & (A3.36||S3.36||G3.36)
13	M1	W3.28 & (L3.29||V3.29||I3.29) & (F2.61||Y2.61||W2.61)
14	M1	N6.52||Q6.52
15	M1	(V6.55||L6.55||I6.55) & (A5.43||V5.43||L5.43||I5.43)
16	AT1	K5.42||R5.42
17	AT1	W6.48 & (F5.47||Y5.47) & (I3.36||V3.36||L3.36||M3.36) &! (W6.52||F6.52||Y6.52)
18	AT1	S3.29||T3.29
19	AT1	V3.32||I3.32||L3.32
20	A2A	N6.55||Q6.55
21	A2A	N6.55||Q6.55
22	A2A	W6.48 & (F6.51||Y6.51||L6.51) & (L3.33)
23	A2A	I7.39||V7.39||L7.39
24	CB1	(F5.42||Y5.42)
25	CB1	W6.48 & (W5.43||Y5.43||F5.43)
26	CB1	(F2.61||Y2.61) & (F2.64||Y2.64)
27	CB1	K3.28
28	CCR2	(E7.39||D7.39) &!K3.32
29	CXCR2	R5.39||K5.39
30	CXCR2	Y6.51
31	CXCR2	(F3.36||W3.36||Y3.36) & (F6.51||Y6.51)
32	CXCR2	E7.39 & K3.32
33	EDG1	E3.29||D3.29
34	EDG1	R3.28||K3.28
35	NAR1	R6.55||K6.55

[a] The required conditions for setting a chemoprint are given (|| = OR; & = AND; &! = AND NOT).

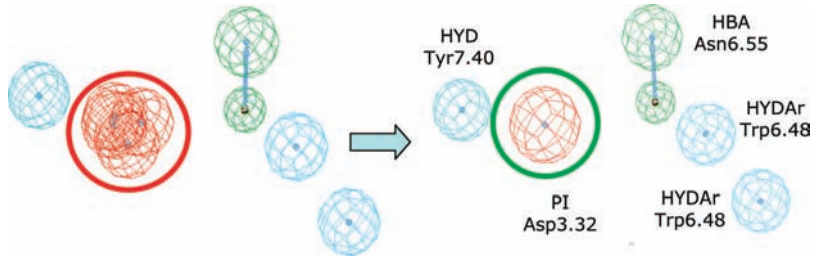


Figure 5. Refinement of initial 3D-pharmacophore model: example, MCH receptor. The sequence-derived 3D-pharmacophore for the MCH receptor is simplified by merging overlapping pharmacophore features of the same type (positively ionizable features). The resulting simplified 3D-pharmacophore is shown on the right.

or any other aliphatic amino acid could recognize a ligand fragment at the same spatial position. As each chemoprint is associated with a single-feature pharmacophore from the pharmacophore building block database, a 3D-pharmacophore can be generated by assembling the appropriate pharmacophore building blocks. The "pharmacophore synthesis" is performed in an automated manner by reading the files of the single-feature pharmacophore building blocks and by combining them into the final 3D-pharmacophore file for the respective receptor. Figure 4 exemplifies the approach for the melanin-concentrating hormone (MCH) receptor.

Automated Refinement of Sequence-Derived Pharmacophores. It is evident from Figure 4 that the resulting sequence-derived pharmacophores might have overlapping pharmacophore features. This is due to the fact that for some chemoprints several single-feature pharmacophores have been stored within the

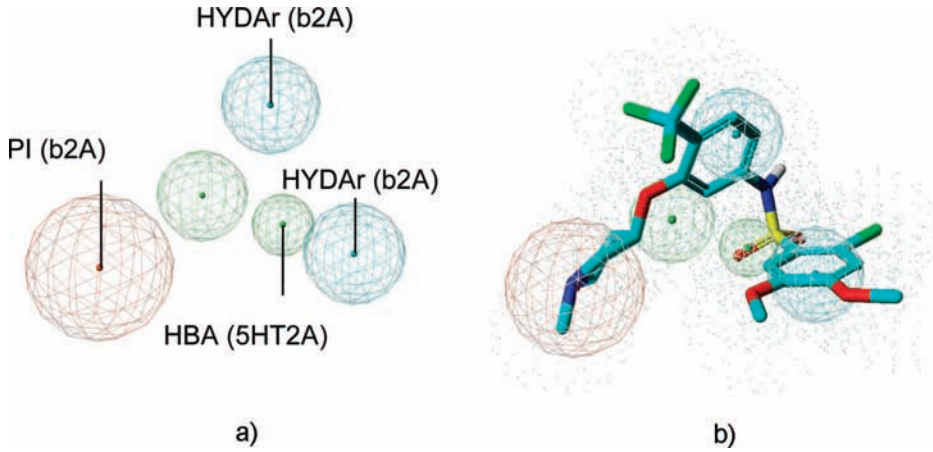


Figure 6. Sequence-derived pharmacophore model for U2 receptor (a) and mapping of a nonpeptidic ligand (**14**) SB706375 (b). The reference ligand has been used to define an additional shape restraint. Each pharmacophore feature is labeled by identity and origin (e.g., the positively ionizable feature has a position as observed in the β2-adrenergic structure).

Table 4. Number of Pharmacophore Features of Sequence-Derived Pharmacophore Models[a]

	<4 features	4 features	>4 features
orphan GPCRs	37	22	34
novel GPCRs	30	12	28

[a] Novel receptors are defined as GPCRs with less than 5 small-molecule ligands in the Aureus GPCR ligand database.[24]

Table 5. Application of Sequence-Derived 3D-Pharmacophore Models in Virtual Screening: Example, U2 Receptor[a]

	U2 receptor ligands retrieved	virtual hits yield (%)	virtual hits from WDI (%)	hit rate (%)
model 1	233	85	6	7
model 2	81	30	0.6	26

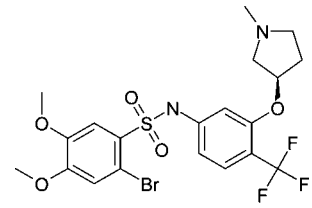
[a] Yield, selectivity (versus WDI), and hit rate are given for the sequence-derived pharmacophore without (model 1) and with shape restraints (model 2).

database with slightly different 3D-coordinates. Another possibility is that a chemoprint signature has been found to make different types of interactions in the given set of reference receptors, e.g., interacting once with a hydrogen bond donor of the ligand and once with a hydrogen bond acceptor (e.g., Asn6.55). This results in overlapping pharmacophore features having different properties. Thus, in order to avoid overlapping features the initial 3D-pharmacophores are cleaned in a second step, again using an automated approach (Figure 5). The initial sequence-derived pharmacophore models of all family A receptors are searched for overlapping pharmacophore features. Identical features are then merged into one feature; overlapping features of a different type are separated and saved as two different 3D-pharmacophore models capturing two different pharmacophore representations of the 7TM binding site of the respective receptor.

By use of the automated approach, sequence-derived 3D-pharmacophore models have been generated for all human family A GPCRs. Table 4 lists the number of pharmacophore features present in the sequence-derived pharmacophore models for orphan GPCRs and novel GPCRs (receptors with less than five known small-molecule ligands within the GPCR ligand database from Aureus).[24] In our experience, 3D-pharmacophore hypotheses with only three pharmacophore features do not provide the necessary "selectivity" when applied in virtual screening, as the number of false positives is too high. However, four- and five-feature hypotheses most often provide a reasonable number of virtual hits when applied in virtual screening of large compound databases. Together with the data shown in Table 4 this suggests that sequence-derived pharmacophore models are applicable for virtual screening of 60% of all orphan GPCRs and for 70% of all novel GPCRs, respectively.

Validation Study: Virtual Screening To Identify Urotensin-II (U2) Receptor Ligands. The sequence-derived pharmacophore model generated for the urotensin-II (U2) receptor is shown in Figure 6a (model 1). The four-feature pharmacophore suggests

Scheme 2



(14) *Ki* 5.4 nM

two hydrophobic aromatic groups, one hydrogen-bond acceptor, and one positively ionizable group in the shown spatial arrangement as common features of small-molecule U2 receptor ligands. How does this sequence-derived 3D-pharmacophore match to the chemical structures of known U2 receptor ligands? The 275 nonpeptidic ligands of the U2 receptor with binding affinities below 1 μM (comprising six different chemical series) have been retrieved from the Aureus database,[24] and a Catalyst database has been generated. The database has been virtually screened using the sequence-derived 3D-pharmacophore model. As a result, 233 hits are retrieved, indicating that 85% of the known nonpeptidic ligands fulfill the pharmacophore requirements captured in the sequence-derived 3D-pharmacophore model.

The applicability of the sequence-derived pharmacophore model (model 1) in virtual screening is evaluated by screening the World Drug Index (WDI) database containing approximately 50 000 druglike molecules combined with the 275 U2 receptor ligands described above. The virtual screening retrieved more than 6% of the database compounds, showing that the selectivity of the four-feature pharmacophore is insufficient for its direct application in virtual screening (Table 5, model 1). The selectivity of a 3D-pharmacophore can be significantly increased

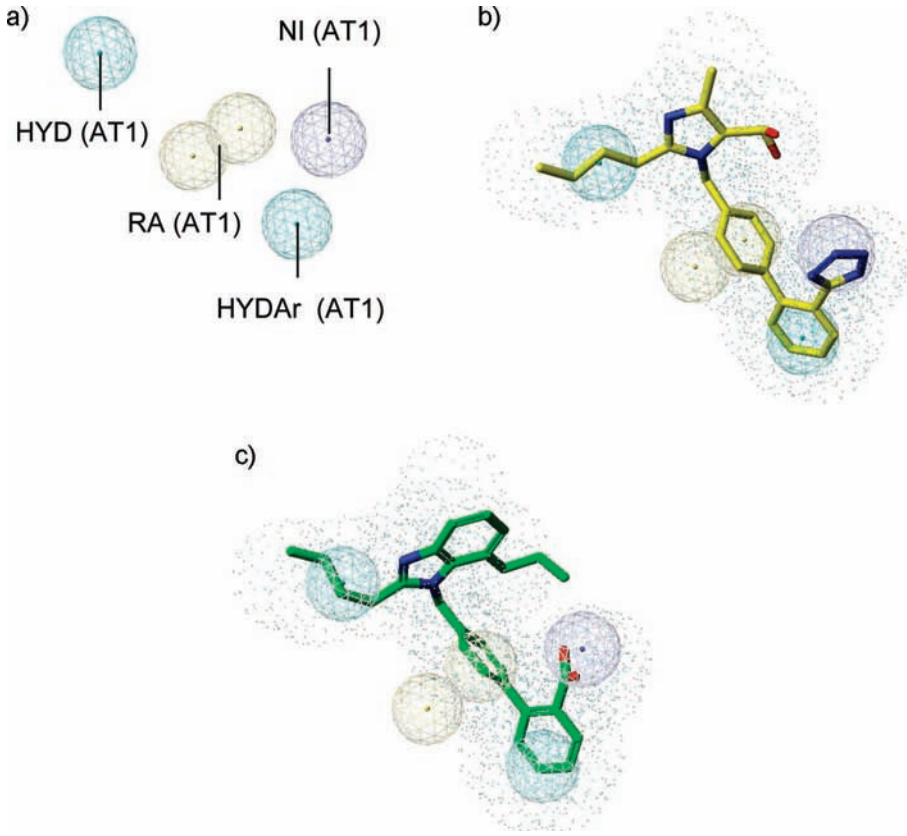
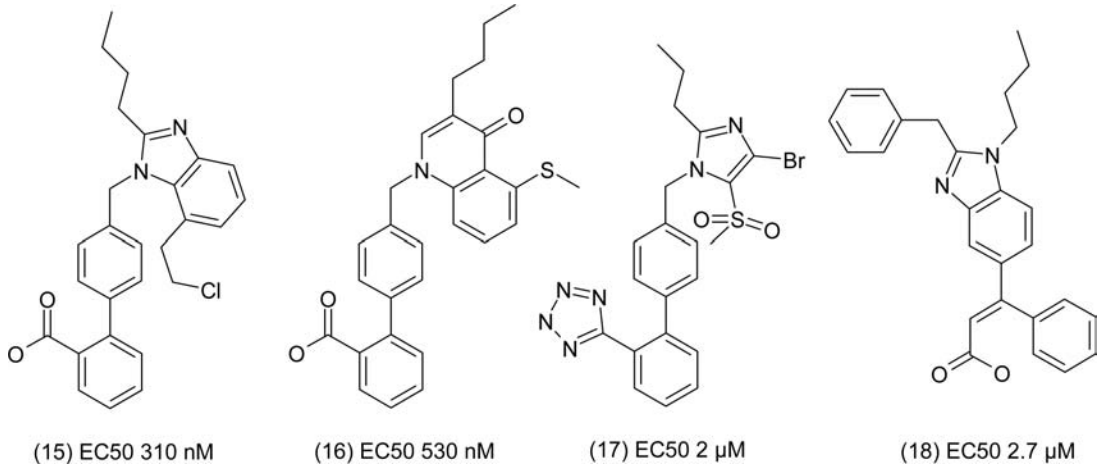


Figure 7. Virtual screening for C3AR1 ligands. (a) The sequence-derived pharmacophore (model 1) consists of four pharmacophore features (ring aromatic, negative ionizable, hydrophobic, and hydrophobic aromatic). For application in virtual screening it is refined by addition of the molecular shape of losartan as additional restraint. (b) The resulting 3D-pharmacophore is applied for virtual screening of the companies' compound database. (c) Mapping of the most potent C3AR1 agonists (**15**) is identified by virtual screening, with an EC_{50} of 300 nM as measured in a functional FLIPR assay.

by adding shape restraints, which can be derived from any potent reference compound. Toward this end, the most active nonpeptidic U2 receptor ligand (**14**, Scheme 2) is mapped onto the sequence-derived 3D-pharmacophore and used as additional shape restraint (Figure 6b, model 2). The shape-refined pharmacophore model (Table 5, model 2) has been applied to virtually screen the WDI database enriched with U2 receptor ligands. The resulting hit list contains 81 (of 275 embedded) U2 receptor ligands (of three different chemical series). Additionally 235 WDI compounds with no reported activity on U2 receptor have been retrieved (putative false positives). Thus, the sequence-derived pharmacophore in combination with shape restraints taken from a reference ligand shows a good selectivity. The pharmacophore retrieves 30% of the known U2 receptor ligands (half of all embedded chemical classes) with a hit rate of 26%. The retrieval rate of active U2 receptor ligands is 50 times higher than a random selection (calculated hit rate of 0.55%). These results from a retrospective virtual screening study encourage the application of sequence-derived pharmacophore models to virtual screening, especially when combined with shape restraints from a reference ligand.

Application Study: Virtual Screening for Ligands of the Complement Component 3a Receptor 1 (C3AR1). In order to support lead finding for the C3AR1, a virtual screening of the companies' compound database has been performed. At the time the virtual screening has been initiated no small molecule ligand of C3AR1 was known. Thus, a sequence-derived 3D-pharmacophore has been generated using the Pharma^{3D} approach described above. The pharmacophore is composed of one ring aromatic feature, a negatively ionizable group, a hydrophobic aromatic group, and a hydrophobic group (Figure 7a). All chemoprints and pharmacophore building blocks originate from the AT1 receptor in complex with losartan (a potent antagonist of the AT1 receptor), indicating a significant similarity between the 7TM binding site of the two peptide-binding GPCRs. This similarity between the C3AR1 and the AT1 receptor is not visible in a classical phylogenic tree based on full sequence homology or by comparison of the receptors based on the physicochemical properties of the amino acids forming the 7TM binding site (using our in-house tool). As the four-feature hypothesis retrieves too many virtual hits when applied for virtual screening of the companies' compound collection, we have added an additional shape restraint. Because of the indicated similarity between C3AR1 and the AT1 receptor, we have used the shape of the losartan molecule, aligned onto the sequence-derived C3AR1 pharmacophore, as additional restraint (Figure 7b). The 650 virtual hits have been retrieved from our companies' collection, and a diverse selection of 157 available compounds has been selected for experimental testing in a functional assay. The compounds have been tested in a fluorometric imaging plate reader (FLIPR) assay using RLB cells expressing the human C3AR1. A total of 4 of the 157 compounds have been found to activate the calcium release with a potency (EC_{50}) below 10 μM (with no activation of the parental RLB cell line lacking the C3AR1). The most potent agonist (mapped onto the pharmacophore model in Figure 7c) shows an EC_{50} of 310 nM (**15**). The potency in the functional assay and the chemical structures of all four C3AR1 agonists

Scheme 3



identified by pharmacophore-based virtual screening are shown in Scheme 3.

Discussion

The Pharma[3D] approach presented here allows the fast generation of sequence-derived 3D-pharmacophore models for GPCRs. The method captures pharmacophore features used in well-studied GPCRs for ligand recognition and extracts the location and identity of pharmacophore features from known GPCR−ligand complexes into a sequence-derived pharmacophore for the GPCR of interest. One retrospective validation study for the human U2 receptor and one prospective application study on C3AR1 demonstrate that these sequence-derived pharmacophore models can be applied in virtual screening. In a retrospective validation study to identify known U2 receptor antagonists the approach provided a hit rate 50 times higher than random screening and identified more than 30% of the U2 ligands hidden in the validation data set. The prospective application study to identify ligands of the C3AR1 by virtual screening of the companies' compound collection and subsequent experimental testing of 157 virtual hits provided four agonists with a potency below 10 μM. The most potent compound (**15**) shows an EC_{50} of 300 nM.

As shown for both application examples, further refinement of the sequence-derived pharmacophore model using the shape of a reference molecule as additional restraint can be beneficial. It increases the selectivity of the model when applied in virtual screening. Alternatively the sequence-derived pharmacophore models can be refined by mapping them onto ligands of the target GPCR resulting, for example, from a low- or medium-throughput screening before application in virtual screening.

Researchers at Roche have recently presented a similar automated approach to derive 3D-pharmacophore models for GPCRs.[25] Kratochwil et al. derive GPCR pharmacophores from coarse homology models placing pharmacophore features next to all amino acids of the 7TM binding site. The resulting 3D-pharmacophores are complex with large coordinate uncertainties for each pharmacophore feature due to the side chain flexibility of the respective interacting amino acid. Thus, according to the authors, these homology-model derived pharmacophore models are found to be useful for the classification of GPCRs but cannot be applied in virtual screening.

Clearly the presented Pharma[3D] approach to derive sequence-derived pharmacophore models for GPCRs has limitations. First, (family A) GPCRs might not all share the same architecture of the 7TM binding site. A similar architecture of the 7TM binding site, however, is a prerequisite for the generation of meaningful pharmacophore models for the target GPCRs using pharmacophore elements of the reference receptors. Second, the current set of 35 pharmacophore building blocks might be insufficient to capture all major interaction sites of all GPCR ligands. Third, interacting residues from the second extracellular loop (ECL2), which can have a contribution to binding (e.g., in the adenosine A2A receptor), are not included in the multiple sequence alignments and are thus currently not considered. All these factors lead to a limited accuracy of the sequence-derived pharmacophore models. We clearly consider the sequence-derived pharmacophore models being of lower quality than 3D-pharmacophore models derived directly from a receptor−ligand cocrystal structure or derived from ligand-based pharmacophore modeling.

After having summarized the limitations of sequence-derived pharmacophores, we emphasize some advantages of this approach. First, the generation of a pharmacophore model in Pharma[3D] is extremely fast. By use of the database of pharmacophore elements and the chemoprint rules as input, 3D-pharmacophore models can be generated in less than 1 min for all family A receptors. Second, especially for GPCRs with limited (or no) ligand information, not allowing the generation of a ligand-based 3D pharmacophore, the sequence-derived pharmacophore models offer the only opportunity to perform pharmacophore-based virtual screening for these targets. Third, the chemoprint information can be used to cluster GPCRs based on common ligand-interaction opportunities. It thus focuses the comparison of binding sites on those residues which are known from other GPCRs to be involved in ligand binding. In addition, common sequence-derived pharmacophore models can be derived addressing a subset of GPCRs and can be applied to direct the design of chemical libraries, targeting a subset of GPCRs (example not shown here). Finally, as seen in the retrospective application example for the U2 receptor, the sequence-derived pharmacophore models often contain elements of different reference receptors (in the U2 receptor study pharmacophore, e.g., elements of the β2-adrenergic receptor and the 5HT2a receptor). Thus, hybrid pharmacophore models are generated that differ from the pharmacophore models of the individual reference receptors. These models thus allow us to address a completely different chemical space when applied in virtual screening compared to searches using the pharmacophore models or reference ligands of the reference receptors as queries.

The Pharma[3D] approach presented here currently relies mainly on homology models and GPCR−ligand complexes supported

by molecular recognition data and SAR data. The reference complexes derived by homology modeling and used as a source for generation of the 3D-pharmacophore building blocks do not have the quality and accuracy of 3D-structures generated by X-ray crystallography. With the expected increasing number of experimental structural data for GPCR−ligand complexes, however, the approach can be based on more experimental structural data. With the data source further improving, the quality and accuracy of the resulting sequence-derived pharmacophore models should also increase in the near future. Thus, we believe that the future of the broad and successful application of the Pharma^{3D} approach for virtual screening and lead finding for novel and orphan GPCRs is still to come.

Acknowledgment. We thank Robert Jäger and Axel Dietrich for uncountable scientific discussions in the area of GPCR chemogenomics, which resulted in the development of Pharma^{3D}. We are also thankful to Markus Mühlbacher, who supported the evaluation of sequence-derived pharmacophore models during an internship at Sanofi-Aventis.

References

(1) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **2002**, *3*, 928–944.
(2) Drug Information Online. www.drugs.com/top200.html (accessed 2007).
(3) Levoye, A.; Jockers, R. Alternative drug discovery approaches for orphan GPCRs. *Drug Discovery Today* **2008**, *13*, 52–58.
(4) Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le, T. 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.
(5) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G.; Thian, F. S.; Kobilka, T. S.; Choi, H. J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
(6) Warne, T.; Serrano-Vega, M. J.; Baker, J. G.; Moukhametzianov, R.; Edwards, P. C.; Henderson, R.; Leslie, A. G.; Tate, C. G.; Schertler, G. F. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature (London)* **2008**, *454*, 486–491.
(7) Jaakola, V. P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.
(8) Schwartz, T. W.; Frimurer, T. M.; Holst, B.; Rosenkilde, M. M.; Elling, C. E. Molecular mechanism of 7TM receptor activation, a global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* **2006**, *46*, 481–519.
(9) Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088–1097.
(10) Evers, A.; Klebe, G. 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) Tikhonova, I. G.; Sum, C. S.; Neumann, S.; Engel, S.; Raaka, B. M.; Costanzi, S.; Gershengorn, M. C. Discovery of novel agonists and antagonists of the free fatty acid receptor 1 (FFAR1) using virtual screening. *J. Med. Chem.* **2008**, *51*, 625–633.
(12) Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2006**, *2*, 207–212.
(13) Kellenberger, E.; Springael, J. Y.; Parmentier, M.; Hachet-Haas, M.; Galzi, J. L.; Rognan, D. Identification of nonpeptide CCR5 receptor agonists by structure-based virtual screening. *J. Med. Chem.* **2007**, *50*, 1294–1303.
(14) Becker, O. M.; Dhanoa, D. S.; Marantz, Y.; Chen, D.; Shacham, S.; Cheruku, S.; Heifetz, A.; Mohanty, P.; Fichman, M.; Sharadendu, A.; Nudelman, R.; Kauffman, M.; Noiman, S. An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist (PRX-00023) for the treatment of anxiety and depression. *J. Med. Chem.* **2006**, *49*, 3116–3135.
(15) Engel, S.; Skoumbourdis, A. P.; Childress, J.; Neumann, S.; Deschamps, J. R.; Thomas, C. J.; Colson, A. O.; Costanzi, S.; Gershengorn, M. C. A virtual screen for diverse ligands: discovery of selective G protein-coupled receptor antagonists. *J. Am. Chem. Soc.* **2008**, *130*, 5115–5123.
(16) Rognan, D. Chemogenomic approaches to rational drug design. *Br. J. Pharmacol.* **2007**, *152*, 38–52.
(17) Klabunde, T. Chemogenomic approaches to drug discovery: similar receptors bind similar ligands. *Br. J. Pharmacol.* **2007**, *152*, 5–7.
(18) Frimurer, T. M.; Ulven, T.; Elling, C. E.; Gerlach, L. O.; Kostenis, E.; Hogberg, T. A physicogenetic method to assign ligand-binding relationships between 7TM receptors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3707–3712.
(19) Evers, A.; Gohlke, H.; Klebe, G. Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *J. Mol. Biol.* **2003**, *334*, 327–345.
(20) Evers, A.; Klebe, G. Ligand-supported homology modeling of G-protein-coupled receptor sites: models sufficient for successful virtual screening. *Angew. Chem., Int. Ed.* **2004**, *43*, 248–251.
(21) Chemical Computing Group, Montreal, Canada. 2006.
(22) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.
(23) *Catalyst*, version 4.11; Accelrys Inc.: San Diego, CA, 2006.
(24) Aureus Pharma. www.aureus-pharma.com (accessed 2004).
(25) Kratochwil, N. A.; Malherbe, P.; Lindemann, L.; Ebeling, M.; Hoener, M. C.; Muhlemann, A.; Porter, R. H.; Stahl, M.; Gerber, P. R. An automated system for the analysis of G protein-coupled receptor transmembrane binding pockets: alignment, receptor-based pharmacophores, and their application. *J. Chem. Inf. Model.* **2005**, *45*, 1324–1336.
(26) Krovat, E. M.; Langer, T. Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. *J. Med. Chem.* **2003**, *46*, 716–726.
(27) 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.
(28) Hamaguchi, N.; True, T. A.; Saussy, D. L., Jr.; Jeffs, P. W. Phenylalanine in the second membrane-spanning domain of alpha 1A-adrenergic receptor determines subtype selectivity of dihydropyridine antagonists. *Biochemistry* **1996**, *35*, 14312–14317.
(29) Hamaguchi, N.; True, T. A.; Goetz, A. S.; Stouffer, M. J.; Lybrand, T. P.; Jeffs, P. W. Alpha 1-adrenergic receptor subtype determinants for 4-piperidyl oxazole antagonists. *Biochemistry* **1998**, *37*, 5730–5737.
(30) Shapiro, D. A.; Kristiansen, K.; Kroeze, W. K.; Roth, B. L. Differential modes of agonist binding to 5-hydroxytryptamine(2A) serotonin receptors revealed by mutation and molecular modeling of conserved residues in transmembrane region 5. *Mol. Pharmacol.* **2000**, *58*, 877–886.
(31) Mitsuya, M.; Kobayashi, K.; Kawakami, K.; Satoh, A.; Ogino, Y.; Kakikawa, T.; Ohtake, N.; Kimura, T.; Hirose, H.; Sato, A.; Numazawa, T.; Hasegawa, T.; Noguchi, K.; Mase, T. A potent, long-acting, orally active (2*R*)-2-[(1*R*)-3,3-difluorocyclopentyl]-2-hydroxy-2-phenylacetamide: novel muscarinic M(3) receptor antagonist with high selectivity for M(3) over M(2) receptors. *J. Med. Chem.* **2000**, *43*, 5017–5029.
(32) Ji, H.; Zheng, W.; Zhang, Y.; Catt, K. J.; Sandberg, K. Genetic transfer of a nonpeptide antagonist binding site to a previously unresponsive angiotensin receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 9240–9244.
(33) McAllister, S. D.; Rizvi, G.; Anavi-Goffer, S.; Hurst, D. P.; Barnett-Norris, J.; Lynch, D. L.; Reggio, P. H.; Abood, M. E. An aromatic microdomain at the cannabinoid CB(1) receptor constitutes an agonist/inverse agonist binding region. *J. Med. Chem.* **2003**, *46*, 5139–5152.
(34) Berkhout, T. A.; Blaney, F. E.; Bridges, A. M.; Cooper, D. G.; Forbes, I. T.; Gribble, A. D.; Groot, P. H.; Hardy, A.; Ife, R. J.; Kaur, R.; Moores, K. E.; Shillito, H.; Willetts, J.; Witherington, J. CCR2: characterization of the antagonist binding site from a combined receptor modeling/mutagenesis approach. *J. Med. Chem.* **2003**, *46*, 4070–4086.
(35) Cutshall, N. S.; Kucera, K. A.; Ursino, R.; Latham, J.; Ihle, N. C. Nicotinanilides as inhibitors of neutrophil chemotaxis. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1517–1520.
(36) Parrill, A. L.; Wang, D.; Bautista, D. L.; Van, B., Jr.; Lorincz, Z.; Fischer, D. J.; Baker, D. L.; Liliom, K.; Spiegel, S.; Tigyi, G. Identification of Edg1 receptor residues that recognize sphingosine 1-phosphate. *J. Biol. Chem.* **2000**, *275*, 39379–39384.
(37) Tunaru, S.; Lattig, J.; Kero, J.; Krause, G.; Offermanns, S. Characterization of determinants of ligand binding to the nicotinic acid receptor GPR109A (HM74A/PUMA-G). *Mol. Pharmacol.* **2005**, *68*, 1271–1280.

JM9001346