

Homology-Based Modelling of Targets for Rational Drug Design

Heather Wieman¹, Kristin Tøndel², Endre Anderssen² and Finn Drabløs^{1*}

¹Department of Cancer Research and Molecular Medicine, Faculty of Medicine, MTFs, Norwegian University of Science and Technology, N-7489 Trondheim, Norway, ²Department of Chemistry, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Abstract: The current status in rational drug design using homology-based models is discussed, with focus on template selection, model building, model verification and strategies for drug design based on model structures. A novel approach for identification of unique binding site features from homology-based models, Protein Alpha Shape Similarity Analysis (PASSA) is described.

Keywords: Homology Model; Drug Design; Template Selection; Model Verification; Protein Alpha Shape Similarity Analysis

INTRODUCTION

Rational drug design is an important concept in pharmaceutical research. The goal is to identify a key drug target based on a thorough understanding of regulatory networks and metabolic pathways, and to design a highly specific drug based on the known three-dimensional (3D) structure of that target. The flood of data from large-scale genome oriented projects is bringing this concept closer to reality. The detailed mapping of genome sequences, regulatory networks and metabolic pathways combined with single nucleotide polymorphism (SNP) data, biological samples or health records makes it easier to identify optimal drug targets. Access to high-quality 3D structures of these targets is a good starting point for rational design of novel drugs.

There are several examples of rational drug design using targets with known 3D structure, including the HIV protease inhibitors amprenavir (Agenerase) and nelfinavir (Viracept) [1-3] and the influenza virus inhibitor zanamivir (Relenza) [4]. Structure-based drug design has also been applied for example in the design of inhibitors of protein kinases [5] such as Abl kinase [6], CDKs [7], EGFR kinase [8], Lck [9] and Src [10].

X-ray crystallography is the main method for structure determination of proteins. This can be a time-consuming process, and it will succeed only if it is possible to find suitable conditions for growing crystals. This can therefore easily become a bottleneck in drug design projects. However, structural domains of proteins can be classified into classes of similar folds, and the number of protein folds actually used by nature seems to be limited [11]. Experimental structure data have been generated for a large fraction of these possible folds, and ongoing structure determination efforts focus on making this mapping as complete as possible. This makes homology-based modelling of protein structures a realistic and relevant alternative to experimental structure determination.

Comparative modelling is often used as a neutral alternative to *homology modelling*, which implies an evolutionary relationship between target and templates. Homology modelling has been used successfully in several drug design projects. Enyedy *et al.* [12] have utilised a homology model of Bcl-2 to identify a novel class of inhibitors by structure-based computer screening. Furet *et al.* [13] successfully applied homology-based modelling for rational design of inhibitors of Cyclin-dependent kinase 1 (CDK1). A modelled structure of an antagonist-bound retinoic acid receptor based on the structure of estrogen receptor has been applied for virtual ligand screening, resulting in the discovery of three novel ligand candidates [14], and homology modelling of Falcipain-2 provided information that led to the discovery of new drug leads against malaria [15].

It is a matter of discussion whether homology models are accurate enough to be utilised in ligand screening and design. It is at least important to use methods that are robust against small structural errors. Recently, Schafferhans and Klebe [16] published a method for computational docking of ligands into protein binding sites that is especially suited for protein structures derived by homology modelling. This method incorporates ligand information into the protein structure modelling procedure. Another drug design method, PASSA, has also been developed specifically for use on homology models. This method uses several alternative homology models for the same protein together with structures of other, related proteins to single out unique features of the target protein [17]. However, such approaches do not decrease the importance of high quality models of potential targets.

HOMOLOGY MODELLING

Homology modelling is based on the observation that the 3D structure of homologous proteins is more conserved than sequence [18]. Chothia and Lesk [19] investigated the relation between sequence conservation and structural similarity for 32 pairs of homologous proteins, and concluded that a protein structure can provide a close general model for other proteins if the sequence similarity is greater than 50%. When the sequence identity drops to 20%, there will be large structural differences. However, the active sites

*Address correspondence to this author at the Department of Cancer Research and Molecular Medicine, Faculty of Medicine, MTFs, Norwegian University of Science and Technology, N-7489 Trondheim, Norway; Tel: +47 73 59 88 42; Fax: +47 73 59 88 01; E-mail: finn.drablos@medisin.ntnu.no

can have very similar geometries, even for distantly related proteins [20,21].

The methodology itself can be described in four steps (illustrated in Fig. (1)): Identifying a suitable template, making an optimal target-template alignment, building the model and validating the model. Protein structure prediction and homology modelling has recently been reviewed by Schonbrun *et al.* [22] and Al-Lazikani *et al.* [23].

Template Identification

The first step is matching the protein sequence of interest (the target) to experimentally determined structures, in order to find at least one protein (the template) for which we can assume that it has the same 3D structure as the target. This is normally based on sequence similarity. Heuristic search methods such as BLAST [24] and FASTA [25] are often used in the initial template-finding step, because these methods are fast and well tested. In difficult cases more sensitive fold recognition methods, which utilise techniques such as Hidden Markov Methods, Neural Networks, iterated searches (e.g. PSI-BLAST [26]), and evolutionary information can be used to scan a structural database for suitable templates [27]. In particular when no close homologues can be found, the increased sensitivity from these methods may allow more potential templates to be identified. This may improve the general reliability of the model, and it may help in identifying structurally conserved regions. For the same reason it is generally advantageous to

use several fold recognition methods in parallel, as alternative algorithms may retrieve slightly different data sets and alignments [28].

Alignment

After identification of the best templates for modelling, an optimal alignment must be made. This seems to be the most crucial step in homology modelling [22,29]. Here "optimal" means that corresponding sequence positions in target and template are identified, so that the predicted structure of the target, based on the template, is as similar as possible to an experimental structure of the same target. Identification of corresponding sequence positions in terms of evolution will at least give a close approximation to an optimal alignment. It is important to realise that the sequence alignment of target with respect to a template identified by a search method or fold recognition method may be sub-optimal with respect to modelling. Different score matrices are needed in order to get optimal alignments for homology modelling as compared to fold recognition [30], possibly because fold recognition needs to focus on conserved regions whereas homology modelling needs to take all regions into account. Hence, alignments generated from fold recognition methods often require refinement in order to be utilised for modelling.

The Smith-Waterman algorithm uses dynamic programming to find an optimal alignment between two sequences, given a scoring matrix and a gap model [31,32].

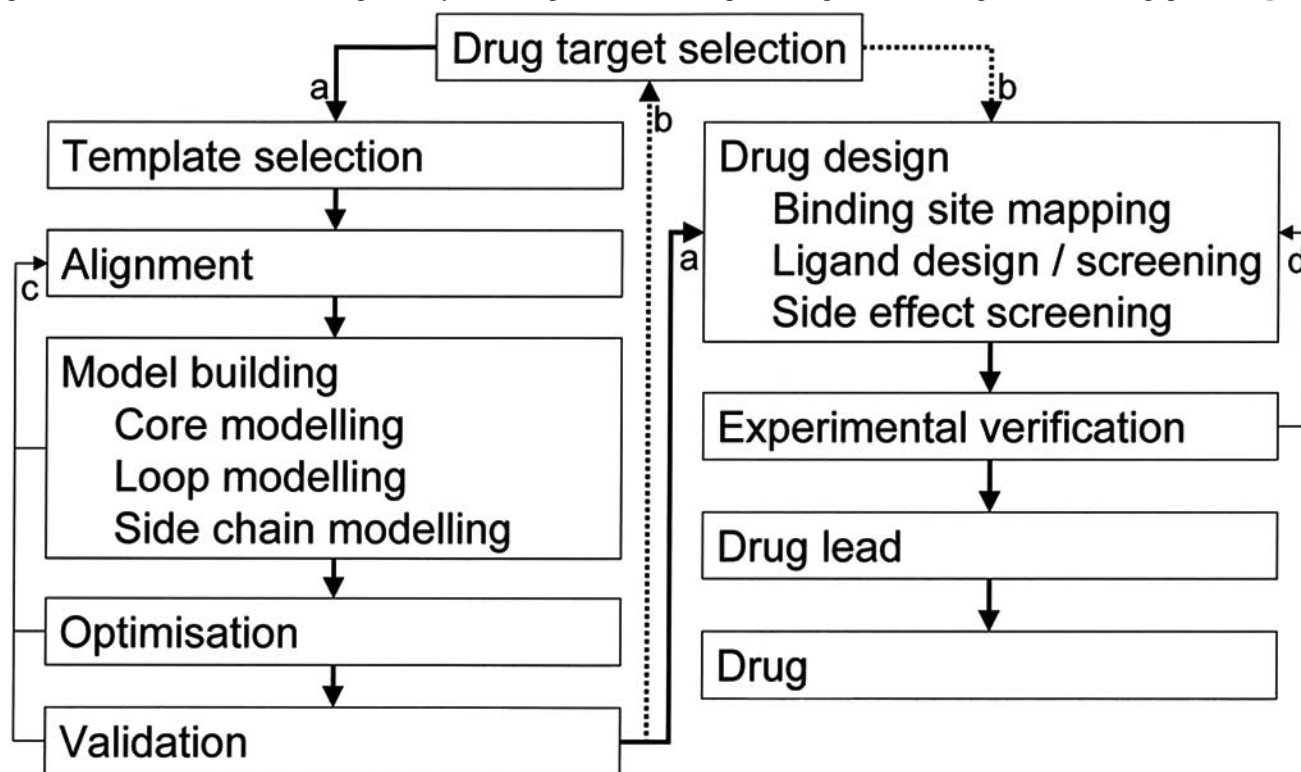


Fig. (1). Key steps in the homology modelling and drug design pathway.

The modelling and design process normally starts with a suitable drug target, the 3D structure of the target is predicted through homology modelling and the model is used for rational drug design (a). However, it is also possible to start with general high-throughput modelling using all potential targets, followed by target selection based on these models (b). In both sections (homology modelling and drug design) there are feedback loops, where e.g. model quality may be used to improve alignment (c) or experimental data may have influence on the drug design strategy and lead optimisation (d).

However, the scoring matrix and gap model represents a simplified model of evolution, and a mathematically optimal solution may still be wrong from an evolutionary perspective. The use of alignments based on multiple sequences is recommended, as this will highlight evolutionary relationships, and increase the probability that corresponding sequence positions are correctly aligned. Moreover, evolutionary information incorporated into sequence profiles greatly increases the alignment accuracy, bringing the alignment closer to the 'true' structural alignment [33]. ClustalX [34], Poa [35], Dialign [36,37] and T-Coffee [38] are important multiple alignment tools. It has been reported that for cases of low sequence identity, Dialign produces the most accurate alignments, whereas T-Coffee is more robust in cases of higher sequence identities [39]. Improved performance can be achieved by combining several alignment strategies [40,41]. Other interesting methods include machine learning [42], fast Fourier transform [43] and improved score matrices built from structural superposition data [44]. New scoring functions have also been developed to give a quantitative measure of alignment accuracy [45].

Using structurally aligned templates as a starting point for the multiple sequence alignment will improve the alignment quality if sequence similarity is low [23,40,46]. Alignment programs such as DALI [47], STRUCTAL [48] and LOCK [49] are examples of structural alignment methods for aligning multiple templates.

Regardless of which program is used, the quality of the alignment should always be verified. However, this is closely related to verification of the homology model itself, and will therefore be discussed there.

Model Building

Model building consists of three main steps. The homology is important mainly when building the core of the protein. Loop modelling is basically *de novo* model building, whereas side chain (re)modelling mainly is an

optimisation step. Reliable identification of structurally conserved core regions versus variable loop regions is an important aspect of this process [50].

There are currently three important approaches for building the core region from alignments. Rigid body superposition constructs the model from a few core sections defined by the average of C α atoms in the conserved regions. Distance geometry uses spatial restraints obtained from the alignment. Segment matching uses a database of short segments of protein structure, with energy or geometry rules, or some combination. It has been shown that when used optimally, accuracies are similar for most modelling methods [51]. Some commonly used programs for homology modelling are listed in Table 1.

SwissModel [52,63] is a popular implementation of the rigid body approach. ProModII [64] generates a model framework based on the topological arrangement of corresponding atoms to the given templates. The backbone is rebuilt based on the positions of C α atoms, using a library of backbone elements derived from high quality X-ray structures. Incomplete loops and incomplete or missing side chains are rebuilt before the models are energy minimised with molecular mechanics (MM).

Homology modelling in MODELLER [56] is based on satisfaction of spatial restraints. Distance and dihedral angle restraints on the target structure are generated, based on the alignment to the template structure. Corresponding distances and angles between aligned residues in the template and the target structures are assumed to be similar. Restraints on bond lengths, bond angles, dihedral angles and nonbonded atom-atom contacts are also derived from statistical analysis of the relationships between C α atoms, solvent accessibilities and side-chain torsion angles in known protein structures. The restraints are expressed as probability density functions (pdfs). These pdfs are combined to give a molecular function, which is optimised using a combination of energy minimisation with molecular dynamics and simulated annealing.

Table 1. Some Commonly Used Homology Modelling Programs

Method	Type ^a	Ref.	Url
SWISS-MODEL	RBS	[52]	http://www.expasy.org/swissmod/SWISS-MODEL.html
WHATIF	RBS	[53]	http://www.cmbi.kun.nl/whatif/
COMPOSER	RBS	[54-58]	http://www.tripos.com/
CONGEN	RBS	[59]	http://www.congenomics.com/
InsightII/Homology	RBS	[60]	http://www.accelrys.com/
TURBO-FRODO	RBS		http://afmb.cnrs-mrs.fr/TURBO_FRODO/
JACKAL	RBS		http://trantor.bioc.columbia.edu/~xiang/jackal
ICM-Homology	RBS	[61]	http://www.molsoft.com/
Look/GeneMine	SM	[62]	http://www.bioinformatics.ucla.edu/genemine/
MODELLER	SR	[56]	http://www.salilab.org/modeller/modeller.html
InsightII/Modeler	SR	[56]	http://www.accelrys.com/

^a RBS – Rigid body superposition, SM – Segment matching, SR – Spatial restraints

The LOOK software package [62] uses Segment Match Modeling (SegMod) to generate homology models by fragment based assembly [65]. SegMod uses a powerful fragment-matching algorithm to find the appropriate structural segments derived from known 3D structures. It utilises both backbone and side chain information from the fragments to obtain a complete model. After building 10 individual models, the averaged model is then minimised using molecular mechanics. SegMod handles insertions and deletions during model building by searching for compatible fragments.

Separate steps are often used for predicting loops (loop libraries or *ab initio* loop building) [66-69], and modelling side chains [68,70-74]. These methods can be used in combination with any of the core modelling techniques.

Although functionally important regions usually are well conserved, flexible loop regions may often contribute significantly in defining specificity. Accurate loop modelling may therefore be important for the usefulness of the homology model. However, existing methods are generally not reliable for loops longer than 5 residues [75]. Loops are often too short to provide sufficient information about their local fold, and segments of up to 9 residues sometimes have entirely unrelated conformations in different proteins [76,77]. Identification of optimal anchor groups seems to be an important step in loop prediction [78,79]. *Ab initio* loop prediction has recently been discussed by Galaktionov *et al.* [69].

After the initial model building, the model can be optimised with molecular mechanics software using either energy minimisation or molecular dynamics methods, or a combination. However, optimisation methods will in general not bring models closer to the true structure [22]. In fact, with extensive refinement homology models actually tend to get worse [80]. Recent data from Flohil *et al.* indicate that some improvement may be gained if long time scale simulation with explicit inclusion of water molecules is used [81]. However, since the roles of optimisation procedures in improving structural quality are still debated [51,68,70], they should be used with caution. Particular care has to be taken when domains, rather than full structures, are modelled. In domain modelling, the positions of any atoms forming an interface to a missing domain should be fixed during energy minimisation. Free movement in these regions can lead to side chain conformations that are preferable energetically, but not possible in the real protein structure because of interactions with residues in the missing domain.

It is still relatively unclear which approach generates the best model. Since 1994 several modelling groups have participated in a bi-annual evaluation project, the Critical Assessment of Techniques for Protein Structure Prediction (CASP) [82]. The groups model proteins that are in the process of being solved experimentally, but not yet have been released for publication. The submitted models are later compared to the then released structures to determine which modelling methods have been most successful. There is also a web-server, EVA-CM (<http://www.pdg.cnb.uam.es/eva/cm/>) which is designed to evaluate protein structure prediction and modelling servers in 'real time' [83,84]. This server evaluates the 'black box' modelling programs. These

programs often limit the number of templates used and impose limitations on manual intervention.

Validation of Models

The accuracy of a model depends upon the sequence similarity it shares with the template. Models with >50% sequence identity to templates are normally of high quality, with ~1 Å root mean square (RMS) error for main chain atoms (equal to medium-resolution NMR or low resolution X-ray structures). Models that have 30 – 50% sequence identity are normally of medium accuracy with an RMS of ~1.5 Å [51,76,85]. Typical errors include problems with side-chain packing, core distortion, loops, and misalignment.

Several validation checks are used for assessing model quality. The most common checks pertain to geometric and stereochemical measurements: covalent geometry (bond lengths and angles), planarity, chirality, phi/psi preferences, chi angles, non-bonded contact distances, unsatisfied donors/acceptors etc [86,87]. Ramachandran plots can provide an overall view of phi/psi values and is a good indicator of the global quality of the model [88]. Quality checks such as these are present in standard crystallographic and NMR software packages as well as in software designed for molecular modelling (e.g. WHATIF and PROCHECK) [53,89]. However, this analysis only indicates the presence of unusual conformations in the structure. Even an incorrect alignment may end up with very reasonable local geometry. Hence, additional tests are needed, in particular for models based on templates with low sequence similarity, where the possibility for misalignment is significant. This is a quite general problem, an interesting example of a misalignment error was recently identified in an experimental 3D structure [90].

Many of these tests are basically fold recognition methods scoring the compatibility between the target sequence and the predicted 3D structure. Sippl *et al.* uses an inverse Boltzmann principle to calculate a mean force potential by 'threading' the target sequence onto structures [91], measuring how well the primary sequence fits the given three-dimensional structure. A related approach tests model correctness by way of a 3-D profile [92]. The 3-D profile of the structure describes the structural environment of each residue. This can be used to score compatibility of any amino acid sequence with that structure. Yet another quality assessment algorithm takes into consideration geometrical parameters of a given structure and then calculates the local, buried and contact energy via statistical potentials of mean force [93-96]. This method has been used in homology modelling to evaluate alternative protein models based on different alignments and as a detector of problematic regions within the protein structure. Another validation measure, designed directly from the results of CASP3, seeks to find the largest subset of C α atoms of the model that can be superpositioned well with the template structures it was modelled from. The normalised score reflects a rough quality measure of the model [97].

The validation checks have to be viewed in light of the validation of the template structure itself. Crystallographic structures are also prone to error, and whatever discrepancies

introduced through the chemical structure determination will most likely also arise in any model based on that structure. The best approach is to gather as much information from as many sources as possible, for both model and templates.

DRUG DESIGN

Given a suitable model of the 3D structure of a potential target, the drug design step tries to find the optimal compound for moderating the normal function of the target in a selective and normally reversible way. In addition to this, several physical criteria have to be met, related to production, uptake, degradation etc. Here we will focus on the actual ligand design, in particular on methods that may improve selectivity. In order to design a ligand for a given target possible interaction sites for ligands have to be identified and the properties of these sites have to be mapped. However, considering only the target protein may be a mistake. Many drugs have recently been withdrawn from late stage testing due to off target effects [98]. Hence, to achieve selectivity and avoid side effects, knowledge of related binding sites is also important. Homology modelling makes this practical, as dozens or hundreds of protein structures can be obtained. If such massive amounts of structural data are to be useful, data analytical methods are needed that aid the interpretation of structural data.

Mapping of Binding Sites

Numerous methods for mapping protein binding sites exist, the majority of which utilise calculations of interaction energies between the protein and small, molecular probes. Binding site analysis is a prerequisite for effective database searches, docking, and *de novo* ligand design. The field of binding site analysis has recently been reviewed [99]. Therefore, this review will focus on mapping strategies that enable comparison of numerous structures for the purpose of understanding selectivity, in particular Multiple Copies Simultaneous Search (MCSS), GRID and Protein Alpha Shape Similarity Analysis (PASSA).

Multiple copies simultaneous search is a method for finding favourable interaction sites in a protein cavity [100]. The idea behind MCSS is to place a large number of copies

of one or more probe molecules into the active site of the target. These probes are placed randomly around the active site atoms and are assumed not to interact with each other (Fig. (2), left). Next, a special energy minimisation protocol is used to refine the initial placement. The receptor atoms may be kept fixed, or be subject to the average forces of the probes [101]. Each probe is subject to the full force of the receptor but not forces from the other probes. Once stable receptor and fragment geometries have been found, fragments with high energies are deleted. The resulting low energy fragments and how they interact with the receptor can then be analysed (Fig. (2), right). The probe molecules are free to move and will have migrated towards regions of favourable interaction with the receptor. This identifies regions of strong interactions that may be used by a ligand. It also gives information on favourable orientation of functional groups. This is useful for *de novo* ligand design as the low energy fragments can be used as starting points. However, a more systematic and complete mapping of the binding site may be necessary, since this random search strategy may not find all relevant interactions.

One of the most common methods for mapping ligand binding sites in proteins is GRID [102], which uses a regular grid spanning the binding site. At each grid point the interaction energy between the protein and a probe group placed on the grid point is computed using a molecular mechanics energy function. Parameters for probes representing various functional groups have been developed [103,104]. The results can be visualised as contour plots of the interaction energies for different probes, and highly detailed potential maps of binding sites may be produced. The low energy contours indicate where functional groups of a ligand are likely to be placed. GRID has been used to suggest functionality for both antibacterial [105,106] and antiviral drugs [107,108]. Both GRID and MCSS have been compared to experimental binding of small molecules by crystallising the same protein in various solvents [109]. It was found that both methods identified approximately the same interaction site, but most results were not reproduced experimentally. In some cases MCSS predicted the correct orientation of the probe, but the predicted orientation of large hydrophobic probes was often wrong. The major reason for the discrepancies between experimental and

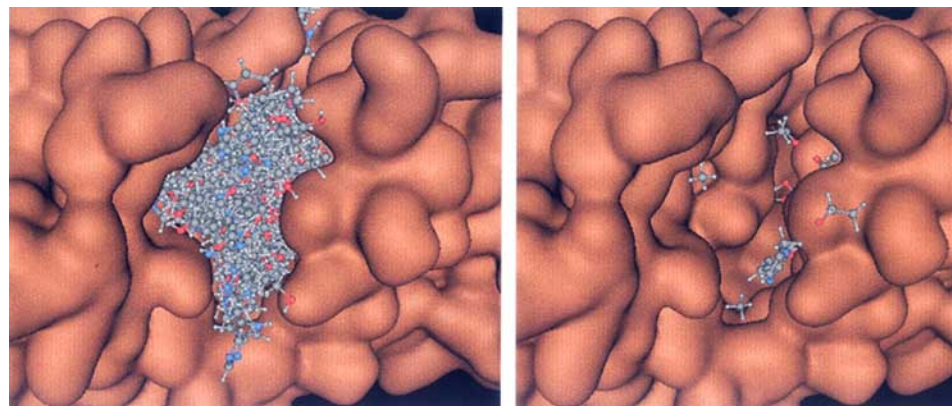


Fig. (2). MCSS.

MCSS mapping of a protein cavity. A large number of different small molecules are placed inside the binding cavity of the protein (left frame). A special energy minimisation procedure is run (see main text). Once the energy minimum has been found, the positions and orientations of low energy molecules may be inspected (right frame).

computational results is believed to be neglect of solvent and entropy effects in the computational models.

Compared to MCSS, the GRID method has the disadvantage that the fragments are not free to move away from their grid point to a more optimal location. However, the fixed grid has a major advantage with respect to comparability. Many related proteins can be superpositioned and the same grid used for all of them. Interesting differences between related binding sites can be identified by comparing energy maps. Thus, using GRID on multiple proteins can aid the development of ligands selective for a

particular protein target. The data from the GRID computations can be analysed by Principal Component Analysis (PCA) to find the most important structural differences to take into consideration for the design of a selective inhibitor. The data analysis tools used to analyse GRID results have been refined by changing the weighting of data from different probes. This has been applied to the design of selective inhibitors of both serine proteases [110] and metalloproteinases [111]. Additional work using homology models has been done on human cytochromes [112].

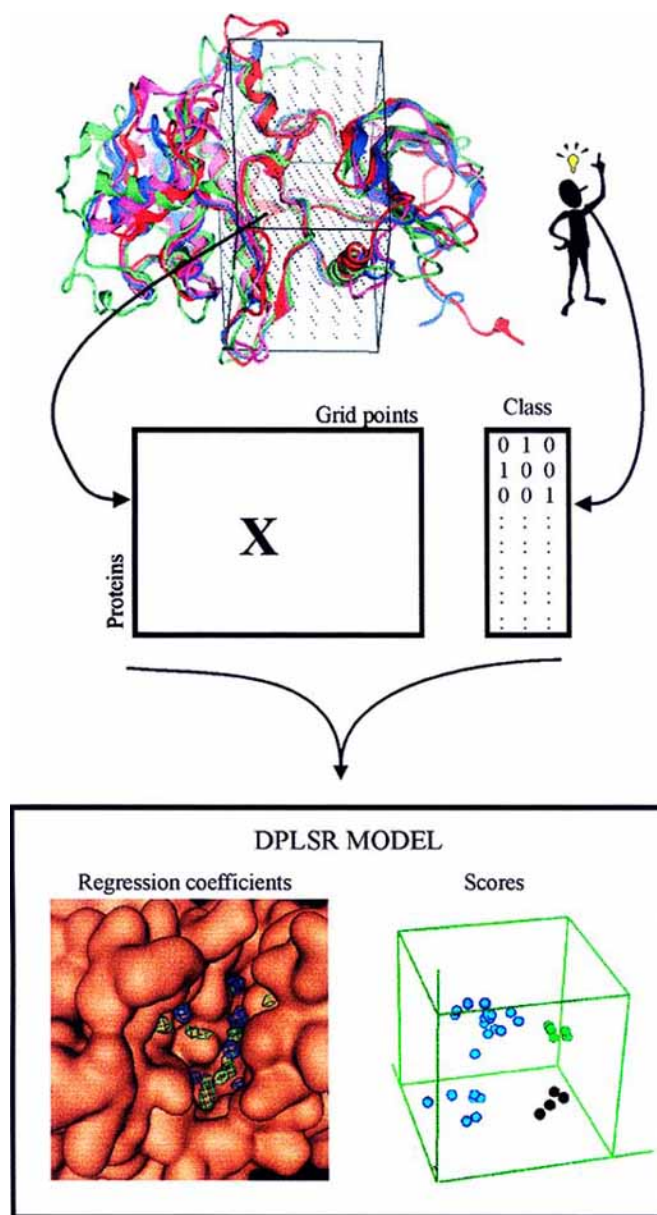


Fig. (3). PASSA.

Protein structures are aligned to maximise the overlap in the active site. A regular grid is placed surrounding the active sites and the alpha shape density of each protein is computed at each grid point. The density data form the matrix X . The user also specifies a number of classes, and assigns each protein to a class. The alpha shape densities and the class data are analysed by DPLSR and a regression model is produced. Interpretation of the model consists of two parts: Regression coefficients and scores. Mapping the regression coefficients back on a protein structure may indicate which regions may contribute to selectivity. Regions may be colour coded by their lipophilic or hydrophilic nature. The scores provide an alternative picture of the model. In the scores space, every protein is represented by one point. Visualising the distribution of proteins in the scores space is useful for discovering clusters, highly deviant structures and structural diversity in a set of proteins.

The use of homology models as the basis for GRID calculations requires some special considerations, specifically if multiple models are to be used in the design of selective ligands. Computing the interaction energy requires precisely defined atomic charges for all atoms, protonation states, and correct placement of hydrogen atoms. The very steep gradients of most force fields close to nuclei may cause instabilities in the PCA models and inflate the effects of small errors in the homology models or the superpositioning.

Protein Alpha Shape Similarity Analysis (PASSA) is an alternative to GRID, developed particularly for use with homology models in the design of selective ligands. This method uses geometrical objects known as alpha spheres to construct a representation of the active site. An alpha sphere is a sphere that contacts four atoms on its surface and has no atoms contained internally. Small alpha spheres correspond to densely packed regions in the protein, while very large spheres are found on the protein surface. In the typical binding pocket however, medium sized spheres are found. Clusters of medium sized spheres will thus correspond to the binding cavities of the protein. Alpha spheres have proven useful for identifying the binding pockets in a number of proteins, and the centres of alpha spheres have been found to correspond well with the placement of atoms in bound ligands [113]. Alpha shapes are determined geometrically, using only the positions and radii of the heavy atoms. This eliminates the need for placing hydrogens and determining protonation states and partial charges. The alpha spheres are classified as hydrophobic or hydrophilic depending on the protein atoms they contact.

PASSA converts the discreet information contained in the placement of alpha sphere centres and protein atoms to a continuous field using a gaussian density estimate. "Dummy" atoms placed at each alpha sphere centre are assigned weights for either the hydrophobic or the hydrophilic field, according to the alpha sphere class. The use of gaussian functions with a very simple partitioning according to the hydrophilic or hydrophobic nature of the alpha spheres reduces some of the problems associated with traditional force field models. Gaussian functions have neither steep derivatives nor singularities. The less detailed representation may also be more robust against the errors typically present in homology models. Analysis of data from gaussian fields typically produce contour plots that are less fragmented and easier to interpret than those produced using force field models [114].

PASSA has been used to suggest properties of a selective inhibitor of Tyrosine kinase 2 (TYK2) and also to understand the basis of the selectivity of STI571, a selective Abl kinase inhibitor [17]. In this approach, Discriminant Partial Least Squares Regression (DPLSR), rather than PCA, is used to analyse the field data (Fig. (3)). DPLSR enables the user to guide the analysis towards features relevant for selectivity towards a specific protein or group of proteins. This is done by dividing the protein structures included in the analysis into classes, typically a 'target' class, containing the structures one wishes to develop a ligand for, and an 'other' class. The 'other' class contains proteins related to the target, but for which a low affinity is desired. Any class scheme may be used e.g. in exploratory

work looking for a suitable drug design target. In some cases, a single protein structure may even belong to more than one class. When analysing homology models in this manner, it is advantageous to use more than one model of each protein, particularly if several templates of comparable sequence identity are available. If several independent structures exist in both the 'target' and 'other' classes, cross validation of the DPLSR model can be used to assess the stability of the model parameters. Thus, the influence of errors in the homology modelling may be gauged. DPLSR works by extracting a low dimensional subspace from the PASSA data that can explain the class structure. Typically relatively few dimensions are needed to separate the classes. This enables visualisation of the relationship between the structure models and easy discovery of clusters or deviant structures. DPLSR models can represent the differences between the protein(s) of interest and all other proteins in the study as a single vector of beta coefficients. The beta coefficients can be visualised as contours in the original 3D space of the protein structures. Spatial regions that may form the basis of selectivity may thus be identified. When designing a TYK2 inhibitor, PASSA was used in combination with MCSS. The plots of the regression coefficients from PASSA were used to guide the selection of MCSS fragments towards those fragments that may contribute to selectivity as well as affinity. This use of combined knowledge of affinity and selectivity is a good starting point for both database searches and *de novo* ligand design, simplifying the task of designing a selective inhibitor.

Database Screening

Once possible interaction sites for a selective inhibitor have been identified, databases of already existing drugs can be searched in order to find a drug molecule that fits the receptor binding site [115]. A number of such databases exist, such as The Cambridge Structural Database [116], the database of The National Cancer Institute (<http://cactus.nci.nih.gov/>), the Available Chemicals Directory (MDL Information Systems) and PDBsum (which includes a database of ligands from the RCSB Protein Database) (<http://www.biochem.ucl.ac.uk/bsm/pdbsum/>). The hits from the database searching can then be evaluated further by molecular docking. Available docking programs include AutoDock [117], DOCK [118], FlexX [119], GOLD [120], LUDI [121] and MOE-Dock (Chemical Computing Group Inc). A version of FlexX suited for combinatorial library docking, FlexX^c, has also been developed [122]. Recently, new docking methods especially suited for use with homology modelled protein structures have been developed. Schafferhans and Klebe [16] use gaussian functions to represent the physico-chemical properties of the receptor and the ligand, and optimise the overlap between the functional description of the receptor binding site and the ligand. Another docking method that utilises gaussian functions is the method developed by McGann *et al.* [123], that acts as a filter to reduce the search space for other docking methods. This method only accounts for shape, and minimises steric clashes between the receptor and ligand atoms. The method developed by Wojciechowski and Skolnick [124] uses a discretisation of the structural models together with an averaging of the structural details and a

smoothing of the potential energy surface to compensate for structural errors. Both steric and chemical complementarity between the ligand and the receptor is sought using a grid-based search. A complete cover of existing docking and virtual screening methods is outside the scope of this review, but the topic has recently been reviewed e.g. by Taylor *et al.* [125], Lyne [126] and Bajorath [127].

To limit the computational time, docking simulations have traditionally been carried out with a fixed protein structure. When using protein structure models built by homology modelling, it is especially important to allow for protein flexibility, since this can reduce the impact of small structural errors. Homology models are built using X-ray structures of other proteins as templates. These are often co-crystallised with a ligand, which induces ligand-specific conformational changes in the protein. Using a rigid protein structure might thus prevent us from identifying optimal binding modes for alternative ligands. Some methods, such as the method developed by Leach [128] and the "Mining Minima Optimizer" method developed by Kairys and Gilson [129] use side-chain flexibility. Anderson *et al.* [130] developed an algorithm for identifying regions where conformational adaptation to a ligand is likely to occur. During the docking simulations the side-chains of these residues are allowed to move.

Recently, some new methods have been developed, that take protein backbone flexibility into account. The majority of these methods utilise multiple protein structure models in the calculations. Österberg *et al.* [131] incorporated protein flexibility and structural water heterogeneity into the

docking simulations using an ensemble of protein structures. In the "Relaxed Complex Method", developed by Lin *et al.* [132,133] a long molecular dynamics (MD) simulation of the unliganded receptor is carried out, followed by a rapid docking of candidate ligands to a large ensemble of the receptor's MD conformations. The FlexE approach [134] is based on a united protein description generated from an ensemble of protein structures. For varying parts of the protein, discrete alternative conformations are explicitly taken into account, which can be combinatorially joined to create new protein structures. Broughton combined the use of statistical analysis of conformational samples from short-run protein molecular dynamics with grid-based docking [135].

De Novo Ligand Design

If one fails to find a drug molecule having the required interacting groups by database searching, the alternative may be to construct a ligand having active groups placed in such a way that interaction with the protein at the identified interaction sites is possible. This ligand construction process is called *de novo* ligand design. A large number of *de novo* design programs are available. These can be divided into three main categories: those that connect molecular fragments placed at the interaction sites to obtain a ligand (linking), those that start from one fragment and connect fragments sequentially to it (growing) and random connection methods. The last category includes the genetic algorithm methods. Most of the random connection methods start from an initial "pool" of fragments and construct

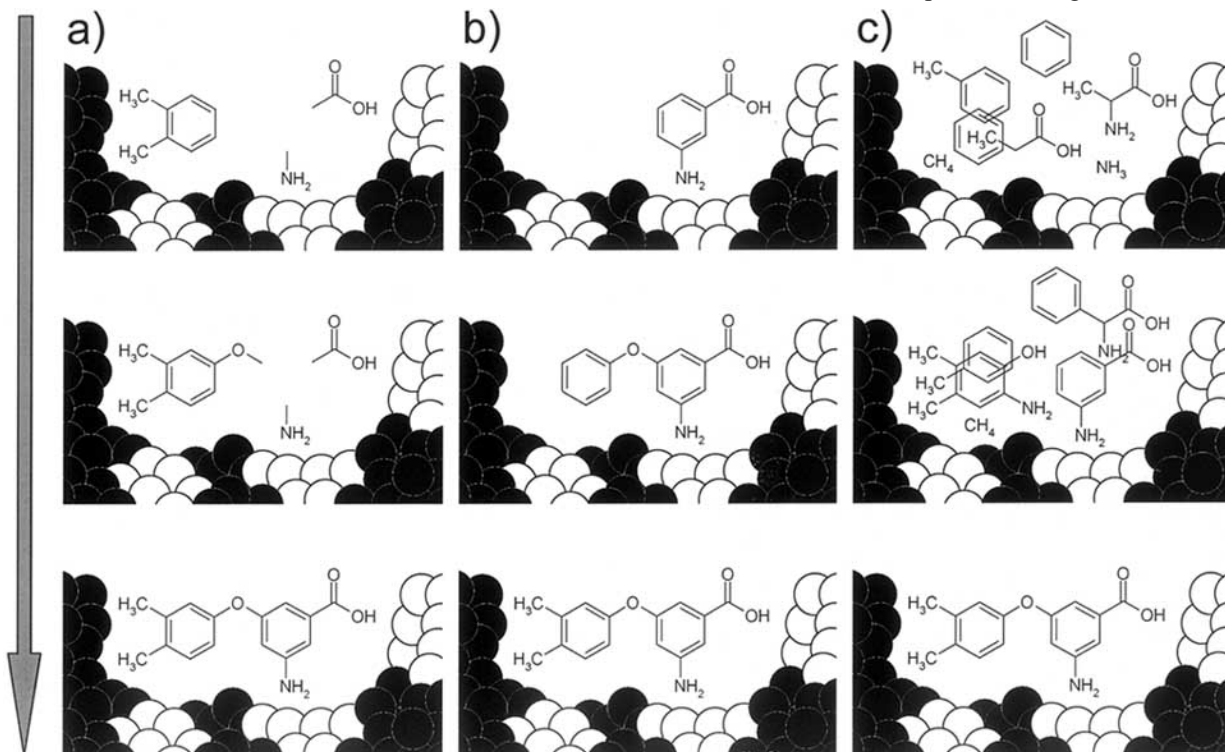


Fig. (4). Three main categories of *de novo* ligand design methods.

Black spheres indicate hydrophobic areas of the protein, while white spheres indicate hydrophilic areas. In the linking approach (a), molecular fragments placed close to important residues of the protein are connected to obtain a ligand. The growing approach (b) starts from one fragment and connects fragments sequentially to it. Most of the random connection methods (c) start from an initial "pool" of fragments and construct ligands by making and breaking connections between the fragments.

Table 2. Some De Novo Ligand Design Programs

Method	Type ^b	Ref	Url
BUILDER	L	[137]	http://thalassa.ca.sandia.gov/~dcroe/
CAVEAT	L	[138]	http://www.cchem.berkeley.edu/~pabgrp/Data/caveat.html
HOOK	L	[139]	http://www.accelrys.com/quanta/mcss_hook.html
LUDI	L	[121]	http://www.accelrys.com/insight/ludi.html
PRO_SELECT	L	[140]	http://www.protherics.com/wtech_camdt.html
SKELGEN	L	[141]	http://www.denovopharma.com/
SmoG	G	[142]	http://www-shakh.harvard.edu/~smog/
CombiSMoG	G	[143]	http://www.concurrentpharma.com/
SPLICE	L	[144]	http://www.tripos.com/
SPROUT	G	[145]	http://www.simbiosys.ca/sprout/
LigBuilder	L+G	[146]	http://mdl.ipc.pku.edu.cn/drug_design/work/ligbuilder.html
LeapFrog	G	Tripos	http://www.tripos.com/
DycoBlock	L	[147]	yyshi@iris.bio.ustc.edu.cn
ADAPT	R	[148]	http://mako.cgl.ucsf.edu/~spegg/
LEA	R	[149]	douguetl@caramail.com

^b L – linking approach, G – growing approach, R – random connection approach

ligands by making and breaking connections between the fragments. Molecular fragments placed at possible interaction sites in the receptor binding pocket found by methods such as PASSA can be used as starting points for all three approaches. These approaches are illustrated in Fig. (4), and Table 2 lists some *de novo* ligand design programs and the approaches they use. A more complete listing of available *de novo* ligand design methods can be found in Schneider *et al.* [136].

There are a number of limitations to existing *de novo* ligand design methods. Most of these methods do not take factors such as synthetic accessibility, bioavailability and metabolic properties into account. Many of the ligand suggestions have large and complex structures. Recently, some programs have been developed that attempt to take such factors into account. An example is LigBuilder [146], which uses a filter to make sure that the structures produced have reasonable ADMET (Absorption, Distribution, Metabolism, Excretion and Toxicity) properties. As for molecular docking, most *de novo* ligand design methods use rigid protein structure models. Recently, some methods have been developed that attempt to take protein flexibility into account. A new version of DycoBlock, F-DycoBlock [150] uses multiple-copy stochastic molecular dynamics to account for fluctuations in the protein structure. Carlson *et al.* developed the “Dynamic Pharmacophore Method” [151], that determines pharmacophore models for a large number of MD snapshots. Protein flexibility in drug design has been reviewed by Carlson and McCammon [152,153] and Wong and McCammon [154].

Most *de novo* ligand design methods use simplified scoring functions for the ligand-receptor system to estimate

binding affinity, mainly in order to speed up the calculations. Solvation effects are typically omitted. Energy-based scoring functions use molecular mechanics force fields to estimate the binding energy, while rule-based scoring functions use rules derived from analysis of structural databases. Energy-based scoring functions are slow, and sensitive to errors in the protein structure, atomic charges and protonation states. In the same way as for GRID, the force field scoring methods are often sensitive to small errors in the atomic positions. Rule-based scoring functions are often very simple, and are highly dependent on the amount of structural data used to derive the rules. In spite of these limitations, *de novo* ligand design methods have contributed to the development of several important drug leads [155], and have proved very useful when combined with some expert knowledge in medicinal chemistry. In recent years several cases of successful application of *de novo* ligand design methods have been reported, as described in the introduction. An important example is the discovery of STI-571, which is a selective inhibitor of Abl kinase, and is currently being used as a therapeutic agent against chronic myelogenous leukaemia [6,156]. Other examples include the development of an antifungal agent [157] using LUDI and the design of aspartyl protease inhibitors using a growth type algorithm. The aspartyl protease inhibitors were verified experimentally [158].

CONCLUSIONS

Homology modelling has significant potential as a tool in rational drug design, in particular in high throughput *in silico* screening or simulation approaches. However, although the methods already are very useful, as

demonstrated in several drug design projects, significant improvement is needed before the tools are robust and general enough for large scale use. All aspects discussed in this review may need some improvement, but a few selected areas may benefit from some extra attention. The quality of the final structure depends mainly on the quality of the target-template alignment. Any improvement in alignment protocols will improve the final model. However, there will always be structural differences between target and templates, and these differences have to be identified and compensated for by *ab initio* modelling or by optimisation methods. In particular optimisation methods based on molecular mechanics and dynamics protocols still represent a weak point, although it is reasonable to assume that it should be possible to improve most models by using a good force field and simulation protocol. Finally, protein structures or ligands are not rigid systems, they have a high degree of flexibility, and docking or design methods that are able to take both the flexibility and small structural errors into account may give improved performance. Improvements in these and other areas may finally turn homology-based rational drug design into a really useful tool for the pharmaceutical industry.

ACKNOWLEDGEMENTS

This work has been supported by the Norwegian research council, projects 139617/140 and 138754/432.

ABBREVIATIONS

QSAR	=	Quantitative Structure-Activity Relationship
CoMSIA	=	Comparative Molecular Similarity Indices Analysis. 3D QSAR method using gaussian property distributions.
CoMFA	=	Comparative Molecular Field Analysis. 3D QSAR method using calculations of interaction energies between the ligands and probe atoms placed on a regular grid.
PCA	=	Principal Component Analysis. Statistical data analysis method.
DPLSR	=	Discriminant Partial Least Squares Regression. Regression method where the dependent variables are indicator variables.
GRID	=	Method for analysis of protein binding sites by calculation of interaction energies between the protein and probe atoms placed on a regular grid.
MCSS	=	Multiple Copies Simultaneous Search. Method for analysis of protein binding sites by calculation of interaction energies between the protein and probe molecules placed in the binding site.
PASSA	=	Protein Alpha Shape Similarity Analysis. Method for analysis of protein binding sites using a combination of gaussian property distributions and DPLSR.
ADMET	=	Absorption Distribution Metabolism Excretion Toxicity

REFERENCES

- [1] Greer, J.; Erickson, J.W.; Baldwin, J.J.; Varney, M.D. *J. Med. Chem.*, **1994**, *37*, 1035.
- [2] Cohen, J. *Science*, **1996**, *272*, 1882.
- [3] Wlodawer, A.; Vondrasek, J. *Annu. Rev. Biophys. Biomol. Struct.*, **1998**, *27*, 249.
- [4] Varghese, J.N. *Drug Dev. Res.*, **1999**, *46*, 176.
- [5] Gray, N.S.; Wodicka, L.; Thunnissen, A.M.; Norman, T.C.; Kwon, S.; Espinoza, F.H.; Morgan, D.O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S.H.; Lockhart, D.J.; Schultz, P.G. *Science*, **1998**, *281*, 533.
- [6] Capdeville, R.; Buchdunger, E.; Zimmermann, J.; Matter, A. *Nat. Rev. Drug Discov.*, **2002**, *1*, 493.
- [7] Davies, T.G.; Tunnah, P.; Meijer, L.; Marko, D.; Eisenbrand, G.; Endicott, J.A.; Noble, M.E. *Structure (Camb.)*, **2001**, *9*, 389.
- [8] Ghosh, S.; Liu, X.P.; Zheng, Y.; Uckun, F.M. *Curr. Cancer Drug Targets*, **2001**, *1*, 129.
- [9] Zhu, X.; Kim, J.L.; Newcomb, J.R.; Rose, P.E.; Stover, D.R.; Toledo, L.M.; Zhao, H.; Morgenstern, K.A. *Structure Fold. Des.*, **1999**, *7*, 651.
- [10] Sawyer, T.; Boyce, B.; Dalgarno, D.; Iulicucci, J. *Expert. Opin. Investig. Drugs*, **2001**, *10*, 1327.
- [11] Govindarajan, S.; Recabarren, R.; Goldstein, R.A. *Proteins*, **1999**, *35*, 408.
- [12] Enyedy, I.J.; Ling, Y.; Nacro, K.; Tomita, Y.; Wu, X.; Cao, Y.; Guo, R.; Li, B.; Zhu, X.; Huang, Y.; Long, Y.Q.; Roller, P.P.; Yang, D.; Wang, S. *J. Med. Chem.*, **2001**, *44*, 4313.
- [13] Furet, P.; Zimmermann, J.; Capraro, H.G.; Meyer, T.; Imbach, P. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 403.
- [14] Schapira, M.; Raaka, B.M.; Samuels, H.H.; Abagyan, R. *Proc. Natl. Acad. Sci. U. S. A.*, **2000**, *97*, 1008.
- [15] Sabnis, Y.; Rosenthal, P.J.; Desai, P.; Avery, M.A. *J. Biomol. Struct. Dyn.*, **2002**, *19*, 765.
- [16] Schafferhans, A.; Klebe, G. *J. Mol. Biol.*, **2001**, *307*, 407.
- [17] Tøndel, K.; Anderssen, E.; Drabløs, F. *J. Comput. Aided Mol. Des.*, **2002**, *16*, 831.
- [18] Lesk, A.M.; Chothia, C. *J. Mol. Biol.*, **1980**, *136*, 225.
- [19] Chothia, C.; Lesk, A.M. *EMBO J.*, **1986**, *5*, 823.
- [20] Raha, K.; Wollacott, A.M.; Italia, M.J.; Desjarlais, J.R. *Protein Sci.*, **2000**, *9*, 1106.
- [21] Zvelebil, M.J.; Barton, G.J.; Taylor, W.R.; Sternberg, M.J. *J. Mol. Biol.*, **1987**, *195*, 957.
- [22] Schonbrun, J.; Wedemeyer, W.J.; Baker, D. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 348.
- [23] Al Lazikani, B.; Jung, J.; Xiang, Z.; Honig, B. *Curr. Opin. Chem. Biol.*, **2001**, *5*, 51.
- [24] Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. *J. Mol. Biol.*, **1990**, *215*, 403.
- [25] Pearson, W.R.; Lipman, D.J. *Proc. Natl. Acad. Sci. U. S. A.*, **1988**, *85*, 2444.
- [26] Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. *Nucleic Acids Res.*, **1997**, *25*, 3389.
- [27] Edwards, Y.J.; Cottage, A. *Mol. Biotechnol.*, **2003**, *23*, 139.
- [28] Lundstrom, J.; Rychlewski, L.; Bujnicki, J.; Elofsson, A. *Protein Sci.*, **2001**, *10*, 2354.
- [29] Venclovas, C. *Proteins*, **2001**, *Suppl. 5*, 47.
- [30] Qian, B.; Goldstein, R.A. *Proteins*, **2002**, *48*, 605.
- [31] Needleman, S.B.; Wunsch, C.D. *J. Mol. Biol.*, **1970**, *48*, 443.
- [32] Smith, T.F.; Waterman, M.S. *J. Mol. Biol.*, **1981**, *147*, 195.
- [33] Jaroszewski, L.; Rychlewski, L.; Godzik, A. *Protein Sci.*, **2000**, *9*, 1487.
- [34] Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. *Nucleic Acids Res.*, **1997**, *25*, 4876.
- [35] Lee, C.; Grasso, C.; Sharlow, M.F. *Bioinformatics*, **2002**, *18*, 452.
- [36] Morgenstern, B.; Frech, K.; Dress, A.; Werner, T. *Bioinformatics*, **1998**, *14*, 290.
- [37] Morgenstern, B. *Bioinformatics*, **1999**, *15*, 211.
- [38] Notredame, C.; Higgins, D.G.; Heringa, J. *J. Mol. Biol.*, **2000**, *302*, 205.
- [39] Lassmann, T.; Sonnhammer, E.L. *FEBS Lett.*, **2002**, *529*, 126.
- [40] Elofsson, A. *Proteins*, **2002**, *46*, 330.
- [41] Lambert, C.; Leonard, N.; De, B.; X; Depiereux, E. *Bioinformatics*, **2002**, *18*, 1250.
- [42] Karwath, A.; King, R.D. *BMC Bioinformatics*, **2002**, *3*, 11.
- [43] Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. *Nucleic Acids Res.*, **2002**, *30*, 3059.

- [44] Blake, J.D.; Cohen, F.E. *J. Mol. Biol.*, **2001**, *307*, 721.
- [45] Cristobal, S.; Zemla, A.; Fischer, D.; Rychlewski, L.; Elofsson, A. *BMC Bioinformatics*, **2001**, *2*, 5.
- [46] Yang, A.S. *Bioinformatics*, **2002**, *18*, 1658.
- [47] Holm, L.; Sander, C. *J. Mol. Biol.*, **1993**, *233*, 123.
- [48] Gerstein, M.; Levitt, M. *Protein Sci.*, **1998**, *7*, 445.
- [49] Singh, A.P.; Brutlag, D.L. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **1997**, *5*, 284.
- [50] Deane, C.M.; Kaas, Q.; Blundell, T.L. *Bioinformatics*, **2001**, *17*, 541.
- [51] Koehl, P.; Levitt, M. *Nat. Struct. Biol.*, **1999**, *6*, 108.
- [52] Guex, N.; Peitsch, M.C. *Electrophoresis*, **1997**, *18*, 2714.
- [53] Vriend, G. *J. Mol. Graph.*, **1990**, *8*, 52.
- [54] Blundell, T.; Carney, D.; Gardner, S.; Hayes, F.; Howlin, B.; Hubbard, T.; Overington, J.; Singh, D.A.; Sibanda, B.L.; Sutcliffe, M. *Eur. J. Biochem.*, **1988**, *172*, 513.
- [55] Johnson, M.S.; Srinivasan, N.; Sowdhamini, R.; Blundell, T.L. *Crit Rev. Biochem. Mol. Biol.*, **1994**, *29*, 1.
- [56] Sali, A.; Blundell, T.L. *J. Mol. Biol.*, **1993**, *234*, 779.
- [57] Sutcliffe, M.J.; Haneef, I.; Carney, D.; Blundell, T.L. *Protein Eng.*, **1987**, *1*, 377.
- [58] Sutcliffe, M.J.; Hayes, F.R.; Blundell, T.L. *Protein Eng.*, **1987**, *1*, 385.
- [59] Bruccoleri, R.E. *Mol. Sim.*, **1993**, *10*, 151.
- [60] Greer, J. *Proteins*, **1990**, *7*, 317.
- [61] Abagyan, R.; Batalov, S.; Cardozo, T.; Totrov, M.; Webber, J.; Zhou, Y. *Proteins*, **1997**, *Suppl. 1*, 29.
- [62] Levitt, M. *J. Mol. Biol.*, **1992**, *226*, 507.
- [63] Guex, N.; Diemand, A.; Peitsch, M.C. *Trends Biochem. Sci.*, **1999**, *24*, 364.
- [64] Peitsch, M.C. *Biochem. Soc. Trans.*, **1996**, *24*, 274.
- [65] Kolodny, R.; Koehl, P.; Guibas, L.; Levitt, M. *J. Mol. Biol.*, **2002**, *323*, 297.
- [66] Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. *Proteins*, **1995**, *23*, 318.
- [67] de Bakker, P.I.; DePristo, M.A.; Burke, D.F.; Blundell, T.L. *Proteins*, **2003**, *51*, 21.
- [68] Fiser, A.; Do, R.K.; Sali, A. *Protein Sci.*, **2000**, *9*, 1753.
- [69] Galaktionov, S.; Nikiforovich, G.V.; Marshall, G.R. *Biopolymers*, **2001**, *60*, 153.
- [70] Chinea, G.; Padron, G.; Hooft, R.W.; Sander, C.; Vriend, G. *Proteins*, **1995**, *23*, 415.
- [71] Dunbrack, R.L. Jr.; Cohen, F.E. *Protein Sci.*, **1997**, *6*, 1661.
- [72] Dunbrack, R.L. Jr. *Curr. Opin. Struct. Biol.*, **2002**, *12*, 431.
- [73] Liu, Z.; Jiang, L.; Gao, Y.; Liang, S.; Chen, H.; Han, Y.; Lai, L. *Proteins*, **2003**, *50*, 49.
- [74] Yang, J.M.; Tsai, C.H.; Hwang, M.J.; Tsai, H.K.; Hwang, J.K.; Kao, C.Y. *Protein Sci.*, **2002**, *11*, 1897.
- [75] Martin, A.C.; MacArthur, M.W.; Thornton, J.M. *Proteins*, **1997**, *Suppl. 1*, 14.
- [76] Marti-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. *Annu. Rev. Biophys. Biomol. Struct.*, **2000**, *29*, 291.
- [77] Mezei, M. *Protein Eng.*, **1998**, *11*, 411.
- [78] Lessel, U.; Schomburg, D. *Proteins*, **1999**, *37*, 56.
- [79] Wohlfahrt, G.; Hangoc, V.; Schomburg, D. *Proteins*, **2002**, *47*, 370.
- [80] Charifson, P.S. *Practical application of computer-aided drug design*, Marcel Dekker, Inc.: New York, **1997**.
- [81] Flohil, J.A.; Vriend, G.; Berendsen, H.J. *Proteins*, **2002**, *48*, 593.
- [82] Moulton, J.; Fidelis, K.; Zemla, A.; Hubbard, T. *Proteins*, **2001**, *Suppl. 5*, 2.
- [83] Eyrich, V.A.; Marti-Renom, M.A.; Przybylski, D.; Madhusudhan, M.S.; Fiser, A.; Pazos, F.; Valencia, A.; Sali, A.; Rost, B. *Bioinformatics*, **2001**, *17*, 1242.
- [84] Marti-Renom, M.A.; Madhusudhan, M.S.; Fiser, A.; Rost, B.; Sali, A. *Structure (Camb.)*, **2002**, *10*, 435.
- [85] Sanchez, R.; Sali, A. *Proc. Natl. Acad. Sci. U. S. A.*, **1998**, *95*, 13597.
- [86] Dodson, E.J.; Davies, G.J.; Lamzin, V.S.; Murshudov, G.N.; Wilson, K.S. *Structure Fold. Des.*, **1998**, *6*, 685.
- [87] MacArthur, M.W.; Laskowski, R.A.; Thornton, J.M. *Curr. Opin. Struct. Biol.*, **1994**, *4*, 731.
- [88] Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.*, **1963**, *7*, 95.
- [89] Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. *J. Appl. Cryst.*, **1993**, *26*, 283.
- [90] Bujnicki, J.; Rychlewski, L.; Fischer, D. *Bioinformatics*, **2002**, *18*, 1391.
- [91] Sippl, M.J. *Proteins*, **1993**, *17*, 355.
- [92] Luthy, R.; Bowie, J.U.; Eisenberg, D. *Nature*, **1992**, *356*, 83.
- [93] Godzik, A.; Kolinski, A.; Skolnick, J. *Protein Sci.*, **1995**, *4*, 2107.
- [94] Godzik, A. *Structure*, **1996**, *4*, 363.
- [95] Jaroszewski, L.; Pawlowski, K.; Godzik, A. *J. Mol. Model.*, **1998**, *4*, 294.
- [96] Pawlowski, K.; Jaroszewski, L.; Bierzynski, A.; Godzik, A. *Pac. Symp. Biocomput.*, **1997**, 328.
- [97] Siew, N.; Elofsson, A.; Rychlewski, L.; Fischer, D. *Bioinformatics*, **2000**, *16*, 776.
- [98] Smith, A. *Nat. Rev. Drug Discov.*, **2002**, *1*, 3.
- [99] Sotriffer, C.; Klebe, G. *Farmacology*, **2002**, *57*, 243.
- [100] Miranker, A.; Karplus, M. *Proteins*, **1991**, *11*, 29.
- [101] Stultz, C.M.; Karplus, M. *Proteins*, **1999**, *37*, 512.
- [102] Goodford, P.J. *J. Med. Chem.*, **1985**, *28*, 849.
- [103] Wade, R.C.; Clark, K.J.; Goodford, P.J. *J. Med. Chem.*, **1993**, *36*, 140.
- [104] Wade, R.C.; Goodford, P.J. *J. Med. Chem.*, **1993**, *36*, 148.
- [105] Powers, R.A.; Shoichet, B.K. *J. Med. Chem.*, **2002**, *45*, 3222.
- [106] Powers, R.A.; Morandi, F.; Shoichet, B.K. *Structure (Camb.)*, **2002**, *10*, 1013.
- [107] von Itzstein, M.; Wu, W.Y.; Kok, G.B.; Pegg, M.S.; Dyason, J.C.; Jin, B.; Van Phan, T.; Smythe, M.L.; White, H.F.; Oliver, S.W. *Nature*, **1993**, *363*, 418.
- [108] von Itzstein, M.; Dyason, J.C.; Oliver, S.W.; White, H.F.; Wu, W.Y.; Kok, G.B.; Pegg, M.S. *J. Med. Chem.*, **1996**, *39*, 388.
- [109] English, A.C.; Groom, C.R.; Hubbard, R.E. *Protein Eng.*, **2001**, *14*, 47.
- [110] Kastenholz, M.A.; Pastor, M.; Cruciani, G.; Haaksma, E.E.; Fox, T. *J. Med. Chem.*, **2000**, *43*, 3033.
- [111] Matter, H.; Schwab, W. *J. Med. Chem.*, **1999**, *42*, 4506.
- [112] Ridderstrom, M.; Zamora, I.; Fjellstrom, O.; Andersson, T.B. *J. Med. Chem.*, **2001**, *44*, 4072.
- [113] Liang, J.; Edelsbrunner, H.; Woodward, C. *Protein Sci.*, **1998**, *7*, 1884.
- [114] Bohm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.*, **1999**, *42*, 458.
- [115] Miller, M.A. *Nat. Rev. Drug Discov.*, **2002**, *1*, 220.
- [116] Allen, F.H.; Kennard, O.; Taylor, R. *Acc. Chem. Res.*, **1983**, *16*, 146.
- [117] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. *J. Comput. Chem.*, **1998**, *19*, 1639.
- [118] Ewing, T.J.A.; Kuntz, I.D. *J. Comput. Chem.*, **1997**, *18*, 1175.
- [119] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. *J. Mol. Biol.*, **1996**, *261*, 470.
- [120] Jones, G.; Willett, P.; Glen, R.C. *J. Mol. Biol.*, **1995**, *245*, 43.
- [121] Bohm, H.J. *J. Comput. Aided Mol. Des.*, **1992**, *6*, 61.
- [122] Rarey, M.; Lengauer, T. *Pers. Drug Disc. Des.*, **2000**, *20*, 63.
- [123] McGann, M.R.; Almond, H.R.; Nicholls, A.; Grant, J.A.; Brown, F.K. *Biopolymers*, **2003**, *68*, 76.
- [124] Wojciechowski, M.; Skolnick, J. *J. Comput. Chem.*, **2002**, *23*, 189.
- [125] Taylor, R.D.; Jewsbury, P.J.; Essex, J.W. *J. Comput. Aided Mol. Des.*, **2002**, *16*, 151.
- [126] Lyne, P.D. *Drug Discov. Today*, **2002**, *7*, 1047.
- [127] Bajorath, J. *Nat. Rev. Drug Discov.*, **2002**, *1*, 882.
- [128] Leach, A.R. *J. Mol. Biol.*, **1994**, *235*, 345.
- [129] Kairys, V.; Gilson, M.K. *J. Comput. Chem.*, **2002**, *23*, 1656.
- [130] Anderson, A.C.; O'Neil, R.H.; Surti, T.S.; Stroud, R.M. *Chem. Biol.*, **2001**, *8*, 445.
- [131] Osterberg, F.; Morris, G.M.; Sanner, M.F.; Olson, A.J.; Goodsell, D.S. *Proteins*, **2002**, *46*, 34.
- [132] Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. *J. Am. Chem. Soc.*, **2002**, *124*, 5632.
- [133] Lin, J.H.; Perryman, A.L.; Schames, J.R.; McCammon, J.A. *Biopolymers*, **2003**, *68*, 47.
- [134] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. *J. Mol. Biol.*, **2001**, *308*, 377.
- [135] Broughton, H.B. *J. Mol. Graph Model.*, **2000**, *18*, 247.
- [136] Schneider, G.; Bohm, H.J. *Drug Discov. Today*, **2002**, *7*, 64.
- [137] Roe, D.C.; Kuntz, I.D. *J. Comput. Aided Mol. Des.*, **1995**, *9*, 269.
- [138] Lauri, G.; Bartlett, P.A. *J. Comput. Aided Mol. Des.*, **1994**, *8*, 51.
- [139] Eisen, M.B.; Wiley, D.C.; Karplus, M.; Hubbard, R.E. *Proteins*, **1994**, *19*, 199.

- [140] Murray, C.W.; Clark, D.E.; Auton, T.R.; Firth, M.A.; Li, J.; Sykes, R.A.; Waszkowycz, B.; Westhead, D.R.; Young, S.C. *J. Comput. Aided Mol. Des.*, **1997**, *11*, 193.
- [141] Todorov, N.P.; Dean, P.M. *J. Comput. Aided Mol. Des.*, **1997**, *11*, 175.
- [142] DeWitte, R.S.; Ishchenko, A.V.; Shakhnovich, E.I. *J. Am. Chem. Soc.*, **1997**, *119*, 4608.
- [143] Grzybowski, B.A.; Ishchenko, A.V.; Shimada, J.; Shakhnovich, E.I. *Acc. Chem. Res.*, **2002**, *35*, 261.
- [144] Ho, C.M.; Marshall, G.R. *J. Comput. Aided Mol. Des.*, **1993**, *7*, 3.
- [145] Gillet, V.; Johnson, A.P.; Mata, P.; Sike, S.; Williams, P. *J. Comput. Aided Mol. Des.*, **1993**, *7*, 127.
- [146] Wang, R.X.; Gao, Y.; Lai, L.H. *J. Mol. Model.*, **2000**, *6*, 498.
- [147] Liu, H.; Duan, Z.; Luo, Q.; Shi, Y. *Proteins*, **1999**, *36*, 462.
- [148] Pegg, S.C.; Haresco, J.J.; Kuntz, I.D. *J. Comput. Aided Mol. Des.*, **2001**, *15*, 911.
- [149] Douguet, D.; Thoreau, E.; Grassy, G. *J. Comput. Aided Mol. Des.*, **2000**, *14*, 449.
- [150] Zhu, J.; Fan, H.; Liu, H.; Shi, Y. *J. Comput. Aided Mol. Des.*, **2001**, *15*, 979.
- [151] Carlson, H.A.; Masukawa, K.M.; McCammon, J.A. *J. Phys. Chem. A*, **1999**, *103*, 10213.
- [152] Carlson, H.A.; McCammon, J.A. *Mol. Pharmacol.*, **2000**, *57*, 213.
- [153] Carlson, H.A. *Curr. Opin. Chem. Biol.*, **2002**, *6*, 447.
- [154] Wong, C.F.; McCammon, J.A. *Annu. Rev. Pharmacol. Toxicol.*, **2003**, *43*, 31.
- [155] Sawyer, T.K. *Biotechniques*, **2001**, *31*, 1164, 1166.
- [156] Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W.T.; Clarkson, B.; Kuriyan, J. *Science*, **2000**, *289*, 1938.
- [157] Ji, H.; Zhang, W.; Zhang, M.; Kudo, M.; Aoyama, Y.; Yoshida, Y.; Sheng, C.; Song, Y.; Yang, S.; Zhou, Y.; Lu, J.; Zhu, J. *J. Med. Chem.*, **2003**, *46*, 474.
- [158] Ripka, A.S.; Satyshur, K.A.; Bohacek, R.S.; Rich, D.H. *Org. Lett.*, **2001**, *3*, 2309.

Copyright of Mini Reviews in Medicinal Chemistry is the property of Bentham Science Publishers Ltd. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.