Utility of homology models in the drug discovery process

Alexander Hillisch, Luis Felipe Pineda and Rolf Hilgenfeld

Advances in bioinformatics and protein modeling algorithms, in addition to the enormous increase in experimental protein structure information, have aided in the generation of databases that comprise homology models of a significant portion of known genomic protein sequences. Currently, 3D structure information can be generated for up to 56% of all known proteins. However, there is considerable controversy concerning the real value of homology models for drug design. This review provides an overview of the latest developments in this area and includes selected examples of successful applications of the homology modeling technique to pharmaceutically relevant questions. In addition, the strengths and limitations of the application of homology models during all phases of the drug discovery process are discussed.

Alexander Hillisch
Bayer HealthCare AG
Apratherweg 18a
D-42096 Wuppertal
Germany
e-mail: alexander.hillisch@bayerhealthcare.com
Luis Felipe Pineda
Jenapharm GmbH & Co. KG
Otto-Schott-Straße 15
D-07745 Jena
Germany
Rolf Hilgenfeld
University of Lübeck
Institute of Biochemistry
Ratzeburger Allee 160
D-23538 Lübeck
Germany

▼ The majority of drugs available today were discovered either from chance observations or from the screening of synthetic or natural product libraries. The chemical modification of lead compounds, on a trial-and-error basis, typically led to compounds with improved potency, selectivity and bioavailability and reduced toxicity. However, this approach is labor- and time-intensive and researchers in the pharmaceutical industry are constantly developing methods with a view to increasing the efficiency of the drug discovery process [1]. Two directions have evolved from these efforts. The 'random' approach involves the development of HTS assays and the testing of a large number of compounds. Combinatorial chemistry is used to satisfy the need for extensive compound libraries. The 'rational', protein structure-based approach relies on an iterative procedure of the initial determination of the structure of the target protein, followed by the prediction of hypothetical ligands for the target protein from molecular modeling and the subsequent chemical synthesis and biological testing of specific compounds (the structure-based drug design cycle). The rational approach is severely limited to target proteins that are amenable to structure determination. Although the protein data bank (PDB; http://www.rcsb.org/pdb) is growing rapidly (~13 new entries daily), the 3D structure of only 1–2% of all known proteins has as yet been experimentally characterized. However, advances in sequence comparison, fold recognition and protein-modeling algorithms have enabled the partial closure of the so-called 'sequence-structure gap' and the extension of experimental protein structure information to homologous proteins. The quality of these homology models, and thus their applicability to, for example, drug discovery, predominantly depends on the sequence similarity between the protein of known structure (template) and the protein to be modeled (target). Despite the numerous uncertainties that are associated with homology modeling, recent research has shown that this approach can be used to significant advantage in the identification and validation of drug targets, as well as for the identification and optimization of lead compounds. In this review, we will focus on the application of homology models to the drug discovery process.

Homology modeling techniques

Homology, or comparative, modeling uses experimentally determined protein structures to predict the conformation of another protein that has a similar amino acid sequence. The method relies on the observation that in nature the structural conformation of a protein is more highly conserved than its amino acid sequence and that small or medium changes in sequence typically result in only small changes in the 3D structure [2].

Generally, the process of homology modeling involves four steps – fold assignment, sequence alignment, model building and model

1359-6446/04/$ – see front matter ©2004 Elsevier Ltd. All rights reserved. PII: S1359-6446(04)03196-4

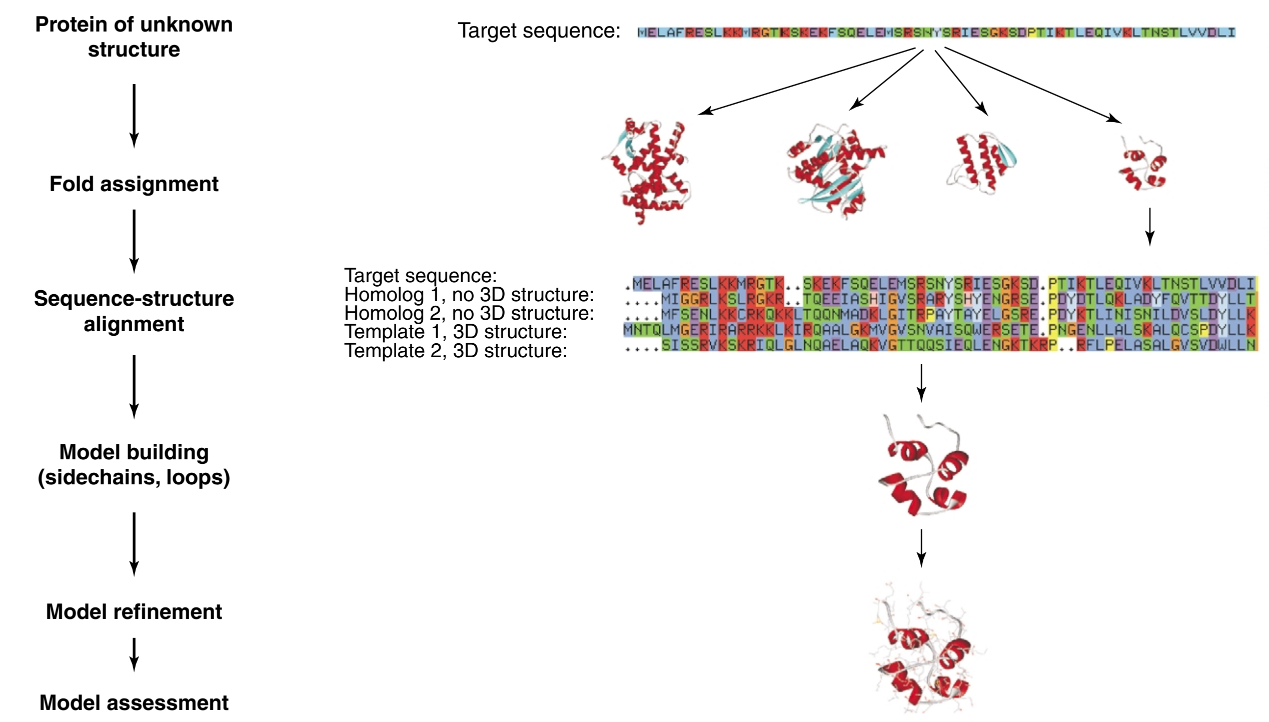

Protein of unknown structure
Fold assignment
Sequence-structure alignment
Model building (sidechains, loops)
Model refinement
Model assessment
Target sequence:
MELAFRESLKKMRGTKSKEKFSQELEMSRSNYSRIESGKSDPTIKTLEQIVKLTNSTLVVDLI
Target sequence:
Homolog 1, no 3D structure:
Homolog 2, no 3D structure:
Template 1, 3D structure:
Template 2, 3D structure:
.MELAFRESLKKMRGTK..SKEKFSQELEMSRSNYSRIESGKSD.PTIKTLEQIVKLTNSTLVVDLI
....MIGGRLKSLRGKR..TQEEIASHIGVSRARYSHYENGRSE.PDYDTLQKLADYFQVTTDYLLT
....MFSENLKKCRKQKKLTQQNMADKLGITRPAYTAYELGSRE.PDYKTLINISNILDVSLDYLLK
MNTQLMGERIRARRKKLKIRQAALGKMVGVSNVAISQWERSETE.PNGENLLALSKALQCSPDYLLK
....SISSRVKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKTKRP..RFLPELASALGVSVDWLLN

Figure 1. The steps involved in the prediction of protein structure by homology modeling. Structure modeling of the bacterial transcriptional repressor CopR is shown [28]. Although the model is based on a low-sequence identity of only 13.8% between CopR and the P22 c2 repressor, several experimental methods support this homology model. Reproduced, with permission, from Ref. [84]. Abbreviation: CopR, plasmid copy control protein.

refinement (Figure 1). The fold assignment process identifies proteins of known 3D structure (template structures) that are related to the polypeptide sequence of unknown structure (the target sequence; this is not to be mistaken with drug target). Next, a sequence database of proteins with known structures (e.g. the PDB-sequence database) is searched with the target sequence using sequence similarity search algorithms or threading techniques [3]. Following identification of a distinct correlation between the target protein and a protein of known 3D structure, the two protein sequences are aligned to identify the optimum correlation between the residues in the template and target sequences. The next stage in the homology modeling process is the model-building phase. Here, a model of the target protein is constructed from the substitution of amino acids in the 3D structure of the template protein and the insertion and/or deletion of amino acids according to the sequence alignment. Finally, the constructed model is checked with regard to conformational aspects and is corrected or energy minimized using force-field approaches.

Several improvements and modifications of this general homology modeling strategy have been developed and applied to the prediction of protein structures. To subject the available structure prediction methods to a blind test, community-wide experiments on the critical assessment of techniques for protein structure prediction (CASP 1–5) have been performed and their results presented and published. As a result, the current state-of-the-art in protein structure prediction has been established, the progress made has been documented and the areas where future efforts might be most productively concentrated have been highlighted [4,5].

Experimental protein structure information and the sequence-structure gap

Homology modeling techniques are dependent on the availability of high-resolution experimental protein structure data. The development of effective protein expression systems and major technological advances in the instrumentation used for structure determination (X-ray crystallography and NMR spectroscopy) has contributed to an exponential growth in the number of experimental protein 3D structures. By May 2004, the PDB contained ~23,000 experimental protein structures for ~7400 different proteins (proteins with less than 90% sequence identity). A recent analysis of all protein chains in the PDB shows that these proteins can be grouped into 2500 protein families comprising 900 unique protein folds [6] (updates can be found at http://scop.mrc-lmb.cam.ac.uk). The majority of the structures in the PDB (84%) were determined by X-ray crystallography, with 15% of the structures being characterized by NMR spectroscopy. The PDB database encompasses experimental information on an extensive array of ligands (small organic molecules and ions) bound to more than 50,000 different binding sites that can be analyzed using programs including ReliBase (http://relibase.ebi.ac.uk) [7], LigBase (http://alto.compbio.ucsf.edu/ligbase) [8] and PDBsum (http://www.biochem.ucl.ac.uk/bsm/pdbsum) [9].

Although the experimental structure database is growing rapidly, there is still a substantial gap between the number of known annotated sequences [1,182,126 unique sequences in Swiss-Prot–TrEMBL (http://www.expasy.org/sprot) as of 29 August 2003] and known protein 3D structures (23,000). If only significantly different proteins are considered (~7400), which omits muteins, artificial proteins and multiple structure determinations of the same proteins (e.g. HIV-protease and carbonic anhydrase II), then less than 1% of the 3D structures of known protein sequences have been elucidated. This sequence-structure gap can partly be filled with homology models. For example, the queryable database ModBase (http://alto.compbio.ucsf.edu/modbase-cgi/index.cgi) provides access to an enormous number of annotated comparative protein structure models [10]. The program PSI-BLAST was used to assign protein folds to all 1,182,126 unique sequence entries in Swiss-Prot–TrEMBL. For 56% of these sequences, comparative models with an average model size of 235 amino acids could be built using the program MODELLER [11]. Thus, by August 2003, 659,495 3D structure models of proteins were accessible via the Internet. The models are predicted to have at least 30% of their C_α atoms superimposed within 3.5 Å of their correct positions. Information on binding sites and ligands can be retrieved from this database using LigBase [8]. However, the majority of the models are built on a low sequence identity and it should be realized that this level of accuracy is, in most cases, not sufficient for a detailed structure-based ligand design.

The SWISS-MODEL Repository (http://swissmodel.expasy.org/repository) [12] is also a database of annotated comparative protein 3D structure models, which have been generated using the fully automated homology-modeling pipeline SWISS-MODEL. As of August 2003, this database contained models for 282,096 different protein sequence entries (26%) from the Swiss-Prot–TrEMBL databases (1,073,566 sequences), with an average model size of ~200 amino acids.

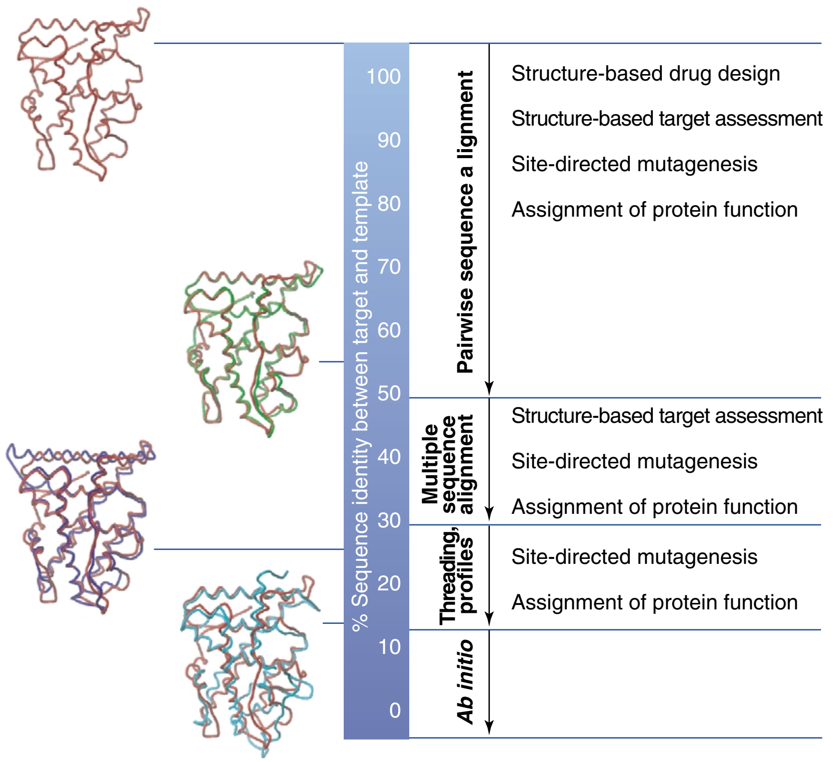
Researchers from Eidogen (http://www.eidogen.com) have created a database system called Target Informatics Platform™ [13] that currently includes homology models for 55,000 proteins. Homology modeling of 26,279 human protein sequences resulted in the construction of 17,442 models for 13,114 different sequences (50%). Thus, putative and known ligand binding pockets can be detected, analyzed and compared and the resulting data used to support target prioritization and lead discovery and/or optimization procedures.

Accelrys (http://www.accelrys.com) produces Discovery Studio (DS) AtlasStore™ as a complete Oracle®-based protein and ligand structural data management solution. Currently, DS AtlasStore™ contains 2,052,000 homology models that have been automatically generated from the sequences of 195,000 proteins from 33 different genomes. In conjunction with homology models, Cengent Therapeutics (http://www.cengent.com) offers dynamic structural information generated from molecular dynamics simulations for 5500 human drug target proteins. This structural information can be used for target prioritization and virtual screening.

Structure information content in homology models

The quality of the homology models is dependent on the level of sequence identity between the protein of known structure and the protein to be modeled [14]. For a sequence identity that is greater than 30%, homology can be assumed; the two proteins probably have a common ancestor and are, therefore, evolutionarily related and are likely to share a common 3D structure. In this case, pairwise and multiple sequence alignment algorithms are reliable and can be used for the generation of homology models (Figure 2).

If the sequence identity is below 15%, structure modeling becomes speculative, which could lead to misleading conclusions. When the sequence identity is between 15% and 30%, conventional alignment methods are not sufficiently reliable and only sophisticated, profile-based methods are capable of recognizing homology and predicting fold. For regions of low sequence identity, threading methods [15] are often applied. Protein models that are built on such low sequence identities can be used for the assignment of protein function and for the direction of mutagenesis experiments (Figure 2). Models that have a sequence identity between ~30% and 50% could facilitate the structure-based prediction of target drugability, the design of mutagenesis experiments and the design of *in vitro* test assays (Figure 2). If sequence identity is greater than ~50%, the resulting models are frequently of sufficient quality to be used in the prediction of detailed protein–ligand interactions, such as structure-based drug design and prediction of the preferred sites of metabolism of small molecules (Figure 2).


100
90
80
70
60
50
40
30
20
10
0
% Sequence identity between target and template
Pairwise sequence a lignment
Multiple sequence alignment
Threading, profiles
Ab initio
Structure-based drug design
Structure-based target assessment
Site-directed mutagenesis
Assignment of protein function
Structure-based target assessment
Site-directed mutagenesis
Assignment of protein function
Site-directed mutagenesis
Assignment of protein function

Drug Discovery Today

Figure 2. Relationship between target and template sequence identity and the information content of resulting homology models. Arrows indicate the methods that can be used to detect sequence similarity between target and template sequences. Applications of the homology models in drug discovery are listed to the right. The higher the sequence identity, the more accurate the resulting structure information. Homology models that are built on sequence identities above ~50% can frequently be used for drug design purposes. Superimpositions of X-ray crystal structures of the ligand-binding domains of members of the nuclear receptor family are shown to the left. These X-ray structures illustrate the increase in structure deviation with a decreased sequence identity. The PR is red, the GR is green, the ERα is blue and the TRβ is cyan. Sequence identities: PR:GR, 54%; PR:ERα, 24%; and PR:TRβ, 15%. Abbreviations: ERα, estrogen receptor α; GR, glucocorticoid receptor; PR, progesterone receptor; TRβ, thyroid receptor β.

Application of homology models in the drug discovery process

There are numerous applications for protein structure information and, hence, homology models at various stages of the drug discovery process [16]. The most spectacular successes are clearly those where protein structural information has helped to identify or to optimize compounds that were subsequently progressed to clinical trials or to the drug market [17]. The applications of homology models that had an impact on target identification and/or validation, lead identification and lead optimization are reviewed here (Figure 3).

Structure-based assessment of target drugability

It is clear that only a minute fraction of the entire proteome can be affected by drug-like (preferentially orally bioavailable) small molecules. Based on the total numbers of known genes, disease-modifying genes and drugable proteins, the number of drug target proteins, for humans, has been estimated at ~600–1500 [18]. For small molecules, sets of properties have been established that differentiate drugs from other compounds [19,20]; these properties can be used to identify compounds with, for example, poor oral absorption properties [21]. Drug molecules and their corresponding target proteins are highly complementary, which suggests that some rules that distinguish good target proteins from others should be deducible [22]. Deep lipophilic pockets that comprise distinct polar interaction sites are clearly superior to shallow highly charged protein surface regions. The inhibition of protein–protein interfaces as a valuable therapeutic principle has recently been shown with inhibitors of the p53–murine double minute clone 2 (MDM2) interaction [23,24]. The binding site for these inhibitors is a distinct lipophilic pocket that normally interacts with the α-helical surface patches of the p53 tumor suppressor transactivation domain. Advances in the rapid detection, description and analysis of ligand-binding pockets [25–27], together with the availability of more than 0.5 million homology models, will open new possibilities for the prioritization of proteins with regards to drugability. In the pharmaceutical industry, structural aspects are being increasingly implemented as additional decision criteria on the drugability of potential drug targets. Companies such as Inpharmatica (http://www.inpharmatica.com) have developed an integrated suite of informatics-based discovery technologies that contain software tools for the structure-based assessment of target drugability.

Structure-guided design of mutagenesis experiments

The design of site-directed mutant proteins is one further important option for the application of homology models to target validation. Introducing point mutations and subsequently studying the effects *in vitro* or *in vivo* is a common approach in molecular biology. This strategy enables the identification of amino acids that are functionally or

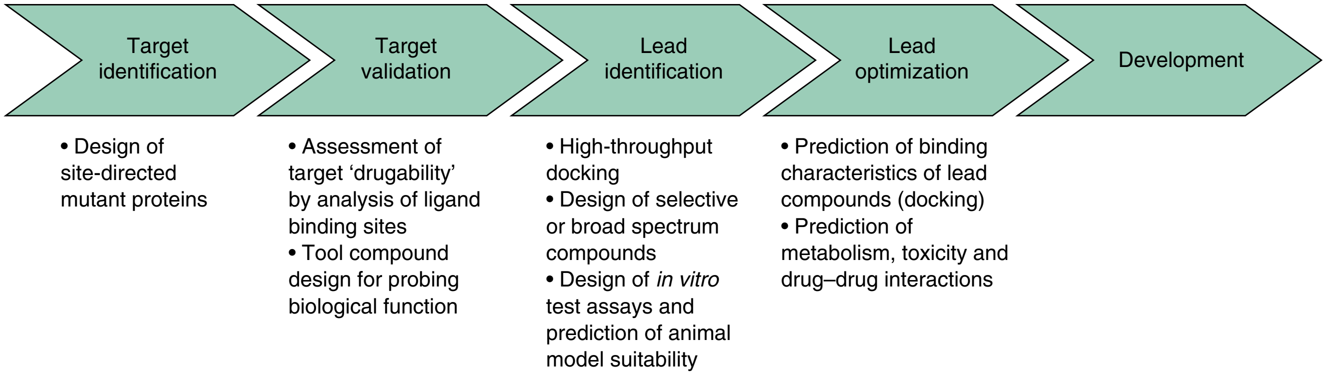

Target identification
Target validation
Lead identification
Lead optimization
Development
• Design of site-directed mutant proteins
• Assessment of target 'drugability' by analysis of ligand binding sites
• Tool compound design for probing biological function
• High-throughput docking
• Design of selective or broad spectrum compounds
• Design of *in vitro* test assays and prediction of animal model suitability
• Prediction of binding characteristics of lead compounds (docking)
• Prediction of metabolism, toxicity and drug–drug interactions

Figure 3. Applications of homology models in the drug discovery process. The enormous amount of protein structure information currently available could not only support lead compound identification and optimization, but could also contribute to target identification and validation. Reproduced, with permission, from [84].

structurally important in the protein under investigation, which ultimately contributes to biological knowledge on, for example, potential target proteins. Typically, the amino acids that are to be modified in these studies are selected on the basis of sequence alignments by focusing on conserved residues. However, if at least some structure information is available, the selection of the amino acids that are to be mutated can be much more precise and successful [28]. This approach is even more powerful when applied in conjunction with pharmacologically active compounds. Site-directed mutants of the target protein can be made to render that target sensitive to an existing pharmacological agent. Based on homology models, some members of the mitogen-activated protein (MAP) kinase family were mutated to make them sensitive to a kinase inhibitor from the pyridinyl imidazole class [29]. This enabled the use of the compound for broader target validation studies.

Tool compound design for probing biological function

One of the most attractive ways to validate a target protein is to administer a pharmacologically active compound that selectively acts on that protein and to study the effects in a relevant animal model. Similar strategies have been described under the term 'chemogenomics' [30].

It has recently been shown that it is possible to design small molecules based on homology models and then to use these compounds as tools to study the physiological role of the respective target protein of that particular drug [31]. Eight years after the discovery of estrogen receptor β (ERβ), the distinct roles of the two ER isotypes, ERα and ERβ, in mediating the physiological responses to estrogens are not completely understood. Although knockout animal experiments have provided an insight into estrogen signaling, additional information on the function of ERα and ERβ was imparted by the application of isotype selective ER agonists. Based on the crystal structure of the ERα-ligand-binding domain (LBD) and a homology model of the ERβ-LBD (59% sequence identity to ERα), Hillisch *et al.* [31] designed steroidal ligands that exploit the differences in size and flexibility of the two ligand-binding cavities (Figure 4). Compounds that were predicted to bind preferentially to either ERα or ERβ were synthesized and tested *in vitro*. This approach led directly to highly ER isotype-selective (200–250-fold) ligands that were also highly potent. To unravel the physiological roles of each of the two receptors, *in vivo* experiments with rats were conducted using the ERα- and ERβ-selective agonists in parallel with the natural ligand of ER, 17β-estradiol. The ERα agonist was shown to be responsible for most of the known estrogenic effects (e.g. induction of uterine growth and bone-protection), in addition to pituitary (e.g. reduction of luteinizing hormone plasma levels) and liver (e.g. increase in angiotensin I plasma levels) effects [31]. However, the ERβ agonist had distinct effects on the ovary, for example, the stimulation of early folliculogenesis [32], which possibly presents clinicians with a new option for tailoring classical ovarian stimulation protocols. A comparison of the homology model with the X-ray crystal structure of the ERβ-LBD complexed with genistein [33] revealed that the homology model had a root-mean-square deviation (rmsd) of the backbone atoms (not considering helix 12) of 1.4 Å. The X-ray crystal structure confirmed the presence of essential interactions between the ligand and the ERβ and did not reveal, at least in this case, any new aspects for the design of ERβ agonists that were not covered by the homology model. These studies show that it is possible to design highly selective compounds, if structure information on all of the relevant homologs of the target is available, and

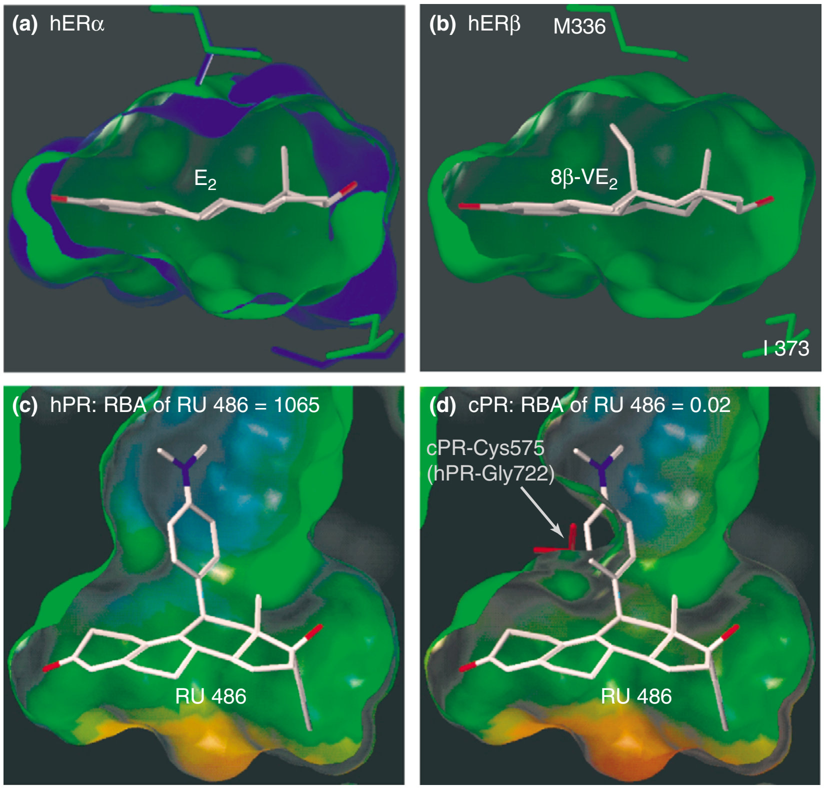


Figure 4. (a,b) Comparison of the two isotypes of the estrogen receptor. In the homology model, ERα (blue) and ERβ (green) ligand-binding pockets are shown in complex with the natural ligand of the ER, 17β-estradiol. The binding of 8β-VE$_2$, a highly potent and selective ERβ agonist, modeled into the ERβ ligand-binding niche is depicted to the right. Reproduced, with permission, from Ref. [31]. **(c,d)** A model of the antiprogestin RU 486 (Mifepristone) bound to hPR. A single amino acid mutation renders this compound inactive at the cPR and hamster PR. Steric clashes between RU 486 and cPR are shown on the right side. Abbreviations: ER, estrogen receptor; hER, human estrogen receptor; cPR, chicken progesterone receptor; hPR, human progesterone receptor; PR, progesterone receptor; RBA, relative binding affinity; 8β-VE$_2$, 8β-vinylestra-1,3,5(10)-triene-3,17β-diol.

that the designed tool compounds contribute to the elucidation of the physiological roles of the target protein.

Homology model-based ligand design

There are numerous examples where protein homology models have supported the discovery and the optimization of lead compounds with respect to potency and selectivity.

Currently, the structures of 40 of the 518 known different human protein kinases have been characterized by X-ray crystallography [34]. Homology model-based drug design has been applied to epidermal growth factor-receptor tyrosine kinase protein [35,36], Bruton's tyrosine kinase [37], Janus kinase 3 [38] and human aurora 1 and 2 kinases [39].

Using the X-ray crystal structure of cyclin-dependent kinase 2 (CDK2), Honma *et al.* [40] generated a homology model of CDK4. This model guided the design of highly potent and selective CDK4 inhibitors that were targeted towards the ATP binding pocket. The diarylurea class of compounds were subsequently synthesized and tested. In an *in vitro* inhibition assay, the most potent compound had an IC_{50} of 42 nM. The predicted binding mode of the lead compound was verified by co-crystallization with CDK2 [40]. Vangrevelinghe *et al.* [41] identified a CDK2 inhibitor using a homology model of the protein and high-throughput docking.

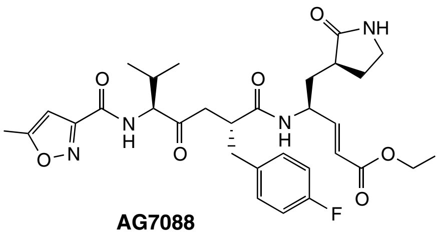
Siedlecki *et al.* [42] have demonstrated the utility of homology modeling in the prediction of pharmacologically active compounds. Alterations in DNA methylation patterns play an important role in tumorigenesis; therefore, inhibitors of DNA methyltransferase 1 (DNMT1), which is the protein that represents the major DNA methyltransferase activity in human cells, are desired. Known inhibitors from the 5-azacytidine class were docked into the active site of a DNMT1 homology model, which led to the design of N4-fluoroacetyl-5-azacytidine derivatives that acted as highly potent inhibitors of DNA methylation *in vitro*.

Thrombin-activatable fibrinolysis inhibitor (TAFI) is an important regulator of fibrinolysis, and inhibitors of this enzyme have potential use in antithrombotic and thrombolytic therapy. Based on a homology model of TAFI (~50% sequence identity to carboxypeptidases A and B), appropriately substituted imidazole acetic acids were designed and were subsequently found to be potent and selective inhibitors of activated TAFI [43].

Homology models of the voltage-gated K^+-channel $K_v1.3$ and the Ca^{2+}-activated channel $IK_{Ca}1$ were used to design selective $IK_{Ca}1$ inhibitors that were based on the polypeptide toxin charybdotoxin. Comparison of the two models revealed a unique cluster of negatively charged residues in the turret of $K_v1.3$ that were not present in $IK_{Ca}1$. To exploit this difference, the homology model was used to design novel analogs, which were then synthesized and tested. Research demonstrated that the novel compounds blocked $IK_{Ca}1$ activity with ~20-fold higher affinity than $K_v1.3$ [44].

The key proteinase (M^pro^, or 3CL^pro^) of the new coronavirus (CoV) that caused the severe acute respiratory syndrome (SARS) outbreak of 2003 (SARS-CoV) is another example of successful homology model building; in this case, success is defined as the ability to use the model to propose an inhibitor that has significant affinity for the target enzyme. X-ray crystal structures for the M^{pro}s of transmissible gastroenteritis virus (TGEV, a porcine coronavirus) and of human coronavirus 229E [45,46] have been characterized. These proteinases have 44 and 40% sequence identity, respectively, with the key proteinase of SARS-CoV. Following publication of the genome sequence of the new virus, first on the internet and a few weeks later in print [46,47], the level of sequence identity between the proteinases enabled Anand *et al.* [46] to construct a 3D homology model for the M^{pro} of human CoV. However, the 3D homology model generated was insufficient for the design of inhibitors with reasonable confidence. To establish the structural basis of the interaction with the polypeptide substrate of the M^{pro}, Anand and co-workers [46] synthesized a substrate-analogous hexapeptidyl chloromethylketone inhibitor that was complexed with TGEV M^{pro}. The X-ray crystal structure of the complex was then determined, which revealed that, as expected, the chloromethylketone moiety had covalently reacted with the active-site cysteine residue of the proteinase. The P1, P2, and P4 side chains of the inhibitor had bound to, and thereby defined, the specificity binding sites of the target enzyme. The experimentally determined structure of the inhibitor–TGEV M^{pro} complex was then compared with all inhibitor complexes of cysteine proteinases in the PDB, which revealed a surprisingly similar inhibitor binding mode in the complex of human rhinovirus type 2 (HRV2) 3C proteinase with AG7088 (Figure 5) [48]. At that time, AG7088 was in late Phase II clinical trials as a drug for the treatment of the strain of the common cold that is caused by human rhinovirus. The comparison of the crystal structures of HRV2 in complex with AG7088 and TGEV M^{pro} in complex with the hexapeptidyl chloromethylketone inhibitor revealed little similarity between the two target enzymes, except in the immediate neighborhood of the catalytic cysteine residue, but an almost perfect match of the inhibitors. To investigate these findings further, AG7088 was docked into the substrate-binding site of the SARS-CoV M^{pro} model without much difficulty, although it was noted that there could potentially be steric problems with the *p*-fluorobenzyl group in the S2 pocket, and also with the ethylester moiety in S1′. Therefore, it was proposed that, although AG7088 was not an ideal inhibitor, this compound should be a good starting point for the design of anti-SARS drugs. Indeed, only a few days after the on-line publication of these results in ScienceXpress [46], it was confirmed that AG7088 had anti-SARS activity *in vitro*. Derivatives of AG7088 with modified P2 residues have since been shown to have K_i values in the lower μmolar range (Rao *et al.*, pers. commun.). The crystal structure of the authentic SARS-CoV key proteinase was determined a few months later [49]. Although the dimeric structure showed the expected similarity to the homologous enzymes of TGEV and human CoV 229E, there were interesting differences in detail. In particular, one of the monomers in the dimer was observed to be in an inactive conformation, which was thought to be the result of the low pH of crystallization. The overall rmsd for the entire dimer from the homology model of Anand *et al.* [46] was >3.0 Å (i.e. no residues excluded from the comparison), which dropped to 2.1 Å when a few outliers at the carboxy terminus were excluded from the comparison, and to <1.8 Å for each of the three individual domains of the enzyme. Other homology models were generated (D. Debe, unpublished and [50]) and virtual screening has been performed using a SARS-CoV M^{pro} model [51]. Taken together, these findings confirm that homology modeling is often inadequate for the prediction of the mutual orientation of domains in multidomain proteins. However, the homology model generated by Anand *et al.* [46] also shows that a reasonable model of a substrate-binding site can serve to develop useful ideas for inhibitor design that can inspire medicinal chemists to start a synthesis program long before the 3D structure of the target enzyme is experimentally determined.

In the case of G-protein-coupled receptors (GPCR), homology-modeling approaches are limited by the lack of experimentally determined structures and the low sequence similarity of those structures that have been characterized with respect to pharmacologically important target proteins. The X-ray crystal structure of only one GPCR, bovine rhodopsin, has been determined [52]. This structure is complemented by bacteriorhodopsin, which is a transmembrane protein that comprises seven helices and is also of relevance for modeling approaches, even though this protein is not a GPCR. Some examples of homology models for GPCRs and their utility have recently been reviewed [53]. High-throughput docking has been applied to verify the ability of homology models to identify agonists (glucocorticoid receptor agonists) [54], antagonists of retinoic acid receptor α [55], D_3-dopamine-, M_1-muscarinic acetylcholine- and V_{1a}-vasopressin-receptors [56] and inhibitors of thrombin [57]. In the identification of thrombin inhibitors, homology models of thrombin were built retrospectively and were based on homologous serine proteases (28%–40% sequence identity); the best docking solutions were yielded with those models that were derived from proteins of higher sequence identity.



AG7088

Figure 5. Structure of AG7088. This compound is an inhibitor of HRV2 3C proteinase and, on the basis of a homology model of HRV2 3C proteinase, was suggested as a potential inhibitor of SARS-CoV M^{pro}. Abbreviations: CoV, coronavirus; HRV2, human rhinovirus type 2; SARS, severe acute respiratory syndrome.

Recently, the performance of docking studies into protein active sites that had been constructed from homology models was assessed using experimental screening datasets of CDK2 and factor VIIa [58]. When the sequence identity between the model and the template near the binding site was greater than ~50%, there was an approximate fivefold increase in the number of active compounds identified than would have been be detected randomly. This performance is comparable to docking to crystal structures.

Design of in vitro *test assays and prediction of animal model suitability*

A further application of homology models is the design of test assays for the *in vitro* pharmacological characterization of compounds or HTS. Based on the structure of the coiled-coil domain of c-Jun, models for α-helical proteins were designed such that they can be used as affinity-tagged proteins that incorporate protease cleavage sites [59]. The resulting 10.5 kDa recombinant proteins were synthesized and used as molecularly defined and uniform substrates for *in vitro* detection of HIV-1 and IgA endoprotease activity, which enabled the surface plasmon resonance-based screening of inhibitors.

The enormous volume of structure information on entire target protein families that is available might also have an impact on screening cascades. Many drug discovery projects endeavor to identify ligands that are highly selective for particular drug targets. Selective compounds are supposed to be superior because such compounds typically lead to fewer adverse side effects (e.g. COX-2 inhibitors). However, the most important homologs that should not be targeted by the desired drug, with respect to the actual target, are not always clear, particularly within the larger target protein families. The sequence similarity of the full-length proteins or entire domains might not always be representative of the target protein when considering the conservation of the ligand-binding pockets. Comparison of the shape and features of the binding pockets within a protein family could indicate which homologs should be included in the screening cascade for so-called 'counter screening'. The structure information that is currently available on entire protein families (e.g. proteases, kinases and nuclear receptors) could contribute to the design of selective compounds or better screening cascades, both of which could potentially advance the design of drugs that have fewer side effects.

A detailed structural knowledge of the ligand-binding sites of target proteins was also shown to facilitate the selection of animal models for *ex vivo* or *in vivo* experiments. The proposal is that animals having target proteins with significantly different binding sites compared with human orthologs should be excluded as pharmacological models. Many promising compounds showing high-potency in human *in vitro* assays have not reached clinical trials because efficacy could not be demonstrated in animal models. Single amino acid differences between humans and animals might, in some cases, be sufficient to cause such effects. The ER selectivity of ligands described by Hillisch *et al.* [31] was shown by homology models and *in vitro* assays to be crucially dependent on the interaction of ligand substituents with one particular amino acid that differs between ERα and ERβ (Figure 4a) [31]. To ensure that this important interaction is present in estrogen receptors of all animal models that are used to characterize compounds [32], homology models of murine, rat and bovine ERβ were built and compared with the binding pocket of human ERβ (hERβ). A complete conservation of amino acids within the binding pockets of human, murine and rat ERβ was observed. However, bovine ERβ showed one amino acid difference at the exact position that was determined to be crucial for ERβ ligand selectivity. The prediction that the hERβ selective compounds should not bind to bovine ERβ was later verified using transactivation experiments (unpublished results). Thus, the implementation of uninterpretable experiments could be avoided at an early stage and the otherwise attractive bovine tissues (later available in larger amounts) could be excluded from *ex vivo* investigations. Similarly, information on the structure of progesterone receptors (PR) can be used to explain the abolished binding of the progesterone antagonist mifepristone (RU 486) to chicken PR and hamster PR [60]. A single point mutation (human PR Gly722 to chicken PR Cys575) prevents antiprogestins containing 11β-aryl substituents (e.g. RU 486) from binding to chicken (and hamster) PR (Figure 4c), which therefore excludes hamsters, for example, from pharmacological studies with antiprogestins [61]. In the future, such effects could be predicted and particular species could then be excluded from pharmacological studies at an early stage, which would ultimately reduce attrition rates in the drug discovery process.

Structure-based prediction of drug metabolism and toxicity

One of the challenges in lead optimization is to identify compounds that not only show a high potency at the desired target protein but also have adequate physical properties to reach systemic circulation, to resist metabolic inactivation for a specific time period and to avoid undesired pharmacological effects. Knowledge of the structure of the proteins that are involved in these processes, such as drug-metabolizing enzymes, transcription factors or transporters, could help to design molecules that do not interact with these 'non-target' proteins.

The cytochrome P450s (CYP) are an extremely important class of enzymes that are involved in Phase I oxidative metabolism of structurally diverse chemicals. Only ~10 hepatic CYPs are responsible for the metabolism of 90% of known drugs. Recently, the X-ray crystal structures of three mammalian CYPs, CYP2C5 [62], CYP2C8 [63] and CYP2C9 [64], have been solved and represent a solid basis for the homology modeling of this entire superfamily. Models of CYP1A2, CYP2A6, CYP2B6, CYP2C8, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP4A11 have been generated using different structure templates. These models have been used to explain and to predict the probable sites of metabolic attack in a variety of CYP substrates [65–72]. However, the large lipophilic and highly flexible character of some CYP binding cavities renders pure *in silico* approaches towards the prediction of the occurrence and site of small molecule metabolism extremely difficult. If protein structure information is combined with pharmacophoric patterns and quantum mechanical calculations, some predictions concerning the preferred sites of metabolism within small molecules are possible [73]. Regarding this aspect of homology modeling, CYP2D6 is a particularly interesting CYP because 5–9% of the Caucasian population does not produce this polymorphic member of the CYP superfamily. The resulting deficiencies in drug oxidation can lead to severe side effects in these individuals. Predictions on whether or not a lead compound could act as a CYP2D6 substrate could help to identify problematic cases early in drug discovery. Combined homology modeling and quantitative SAR approaches are able to predict such CYP inhibitors [74]. Thus, in the future, protein structure information in conjunction with high-throughput docking and pharmacophore-based methods could be used to decide which compounds have the potential to inhibit particular CYPs. This approach could facilitate the detection of potential drug–drug interactions early in the drug discovery process and measures could then be taken to avoid such interactions [75].

CYP substrates and inhibitors are not the only compounds to have been studied using homology models. These approaches have been used recently to investigate CYP inducers. The induction of CYPs is primarily mediated via the activation of ligand-dependent transcription factors, such as the aryl hydrocarbon receptor (AhR) for the CYP1A family, the constitutive androstane receptor (CAR) for the CYP2D family and the pregnane X receptor (PXR), glucocorticoid receptor (GR) and vitamin D receptor (VDR) for the CYP3A family [76]. In principle, the *in silico* prediction of drug-metabolizing enzyme induction could be reduced to predicting the binding and activation of transcription factors (e.g. AhR and CAR). However, recent X-ray structure analyses of PXR have shown that the LBD of this nuclear receptor contains a large lipophilic and flexible binding pocket [77]. This renders pure *in silico* structure-based predictions concerning whether or not a small molecule will activate PXR difficult. The homology modeling of CAR [78,79] and other members of the nuclear receptor family involved in CYP induction [80] have recently been described. These models predict reasonably shaped potential ligand binding pockets. However, further results on the utility of these models are needed.

With respect to the structure-based prediction of adverse health effects, progress has been described with the human ether-a-go-go-related gene (hERG). This tetrameric potassium channel contributes to phase three repolarization of heart muscle cells by opposing the depolarizing Ca^{2+} influx during the plateau phase. Inhibition of this protein results in cardiovascular toxicity (QT-prolongation) and has caused several drugs to be withdrawn from the market. Therefore, *in silico* predictions on the probability of the formation of an interaction between a drug and hERG have gained enormous attention and have recently been reviewed [81]. Homology models of hERG, which are based on the X-ray crystal structures of the bacterial KcsA [82] and MthK channels [83], have already shed light on some details of the molecular interactions that initiate hERG inhibition. However, the complexity of this potassium channel signifies that detailed X-ray structure analyses of the protein in the open- and closed-state are required before these molecular interactions can be fully understood and predicted, which has implications for the prediction of cardiotoxicity.

Conclusions and outlook

Numerous examples for the successful application of homology modeling in drug discovery are described here. In the absence of experimental structures of drug target proteins, homology models have supported the design of several potent pharmacological agents. One of the advantages of homology models is that these models can be generated relatively easily and quickly. Furthermore, such models could support the hypotheses of medicinal chemists on how to generate biologically active compounds in the important early conceptual phase of a drug discovery project. The design of compounds that are selectively directed at particular drug target proteins is one of the strengths of this technique. Such selective compounds can even be applied to gain insights into the physiological role of novel drug targets. The *in silico* protein structure-based prediction of metabolism and toxicity of small molecules, particularly CYP inhibition and induction and hERG inhibition, is currently in its infancy and predictive capabilities could be limited to classification only. However, while complete experimental structures of pharmacologically important proteins are missing, the homology modeling technique provides one approach to bridge the gap until this information becomes available.

Acknowledgements

We gratefully acknowledge fruitful discussions with Mario Lobell (Bayer HealthCare; http://www.bayerhealthcare.com), Derek Debe, Sean Mullen (Eidogen) and Sunil Patel (Accelrys).

References

1 Giersiefen, H. *et al.* (2003) Modern Methods of Drug Discovery: An Introduction. In *Modern Methods of Drug Discovery* (Hillisch, A. and Hilgenfeld, R., eds), pp. 1–18, Birkhäuser Verlag

2 Lesk, A.M. and Chothia, C. (1986) The response of protein structures to amino-acid sequence changes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 317, 345–356

3 Godzik, A. (2003) Fold Recognition Methods. In *Structural Bioinformatics* (Bourne, P. and Weissig, H., eds), pp. 525–546, Wiley-Liss

4 Murzin, A.G. (2001) Progress in protein structure prediction. *Nat. Struct. Biol.* 8, 110–112

5 Tramontano, A. *et al.* (2003) Assessment of homology-based predictions in CASP5. *Proteins* 53, 352–368

6 Murzin, A.G. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540

7 Hendlich, M. (1998) Databases for protein–ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* 54, 1178–1182

8 Stuart, A.C. *et al.* (2002) LigBase: a database of families of aligned ligand binding sites in known protein sequences and structures. *Bioinformatics* 18, 200–201

9 Laskowski, R.A. (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* 29, 221–222

10 Pieper, U. *et al.* (2004) MODBASE, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 32, D217–D222

11 Sali, A. *et al.* (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815

12 Kopp, J. *et al.* (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res.* 32, D230–D234

13 Debe, D.A. and Hambly, K. (2004) Supporting your pipeline with structural knowledge. *Curr. Drug Discov.* 3, 15–18

14 Chothia, C. *et al.* (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J.* 5, 823–826

15 Kinch, L.N. *et al.* (2003) CASP5 assessment of fold recognition target predictions. *Proteins* 53, 395–409

16 Hillisch, A. and Hilgenfeld, R. (2003) The role of protein 3D structures in the drug discovery process. In *Modern Methods of Drug Discovery* (Hillisch, A. and Hilgenfeld, R., eds), pp. 157–181, Birkhäuser Verlag

17 Hardy, L.W. *et al.* (2003) The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.* 12, 15–20

18 Hopkins, A.L. *et al.* (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727–730

19 Walters, W.P. *et al.* (2002) Prediction of ‘drug-likeness’. *Adv. Drug Deliv. Rev.* 54, 255–271

20 Sadowski, J. *et al.* (1998) A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* 41, 3325–3329

21 Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* 44, 235–249

22 Fauman, E.B. *et al.* (2003) Structural bioinformatics in drug discovery. *Methods Biochem. Anal.* 44, 477–497

23 Chene, P. (2004) Inhibition of the p53-hdm2 interaction with low molecular weight compounds. *Cell Cycle* 3, 460–461

24 Chene, P. (2004) Inhibition of the p53-MDM2 interaction: targeting a protein–protein interface. *Mol. Cancer Res.* 2, 20–28

25 Schmitt, S. *et al.* (2002) A new method to detect related function among proteins independent of sequence and fold homology. *J. Mol. Biol.* 323, 387–406

26 Binkowski, T.A. *et al.* (2003) Inferring functional relationships of proteins from local sequence and spatial surface patterns. *J. Mol. Biol.* 332, 505–526

27 Naumann, T. *et al.* (2002) Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* 45, 2366–2378

28 Steinmetzer, K. *et al.* (2000) Transcriptional repressor CopR: structure model-based localization of the deoxyribonucleic acid binding motif. *Proteins* 38, 393–406

29 Eyers, P.A. *et al.* (1999) Use of a drug-resistant mutant of stress-activated protein kinase 2a/p38 to validate the *in vivo* specificity of SB 203580. *FEBS Lett.* 451, 191–196

30 Bredel, M. *et al.* (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* 5, 262–275

31 Hillisch, A. *et al.* (2004) Dissecting physiological roles of estrogen receptor alpha and beta with potent selective ligands from structure-based design. *Mol. Endocrinol.* 18, 1599–1609

32 Hegele-Hartung, C. *et al.* (2004) Impact of isotype-selective estrogen receptor agonists on ovarian function. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5129–5134

33 Pike, A.C. *et al.* (1999) Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *EMBO J.* 18, 4608–4618

34 Manning, G. *et al.* (2002) The protein kinase complement of the human genome. *Science* 298, 1912–1934

35 Traxler, P. *et al.* (1997) Design and synthesis of novel tyrosine kinase inhibitors using a pharmacophore model of the ATP-binding site of the EGF-R. *J. Pharm. Belg.* 52, 88–96

36 Ghosh, S. *et al.* (2001) Rational design of potent and selective EGFR tyrosine kinase inhibitors as anticancer agents. *Curr. Cancer Drug Targets* 1, 129–140

37 Mahajan, S. *et al.* (1999) Rational design and synthesis of a novel anti-leukemic agent targeting Bruton’s tyrosine kinase (BTK), LFM-A13. *J. Biol. Chem.* 274, 9587–9599

38 Sudbeck, E.A. *et al.* (1999) Structure-based design of specific inhibitors of Janus kinase 3 as apoptosis-inducing antileukemic agents. *Clin. Cancer Res.* 5, 1569–1582

39 Vankayalapati, H. *et al.* (2003) Targeting aurora2 kinase in oncogenesis: a structural bioinformatics approach to target validation and rational drug design. *Mol. Cancer Ther.* 2, 283–294

40 Honma, T. *et al.* (2001) Structure-based generation of a new class of

potent Cdk4 inhibitors: new *de novo* design strategy and library design. *J. Med. Chem.* 44, 4615–4627

41 Vangrevelinghe, E. *et al.* (2003) Discovery of a potent and selective protein kinase CK2 inhibitor by high-throughput docking. *J. Med. Chem.* 46, 2656–2662

42 Siedlecki, P. *et al.* (2003) Establishment and functional validation of a structural homology model for human DNA methyltransferase 1. *Biochem. Biophys. Res. Commun.* 306, 558–563

43 Barrow, J.C. *et al.* (2003) Synthesis and evaluation of imidazole acetic acid inhibitors of activated thrombin-activatable fibrinolysis inhibitor as novel antithrombotics. *J. Med. Chem.* 46, 5294–5297

44 Rauer, H. *et al.* (2000) Structure-guided transformation of charybdotoxin yields an analog that selectively targets Ca^{2+}-activated over voltage-gated K^+ channels. *J. Biol. Chem.* 275, 1201–1208

45 Anand, K. *et al.* (2002) Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. *EMBO J.* 21, 3213–3224

46 Anand, K. *et al.* (2003) Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767

47 Rota, P.A. *et al.* (2003) Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300, 1394–1399

48 Matthews, D.A. *et al.* (1999) Structure-assisted design of mechanism-based irreversible inhibitors of human rhinovirus 3C protease with potent antiviral activity against multiple rhinovirus serotypes. *Proc. Natl. Acad. Sci. U. S. A.* 96, 11000–11007

49 Yang, H. *et al.* (2003) The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13190–13195

50 Takeda-Shitaka, M. *et al.* (2004) Evaluation of homology modeling of the severe acute respiratory syndrome (SARS) coronavirus main protease for structure-based drug design. *Chem. Pharm. Bull. (Tokyo)* 52, 643–645

51 Xiong, B. *et al.* (2003) A 3D model of SARS CoV 3CL proteinase and its inhibitors design by virtual screening. *Acta Pharmacol. Sin.* 24, 497–504

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

53 Becker, O.M. *et al.* (2003) Modeling the 3D structure of GPCRs: advances and application to drug discovery. *Curr. Opin. Drug Discov. Devel.* 6, 353–361

54 Schapira, M. *et al.* (2003) Nuclear hormone receptor targeted virtual screening. *J. Med. Chem.* 46, 3045–3059

55 Schapira, M. *et al.* (2000) Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1008–1013

56 Bissantz, C. *et al.* (2003) Protein-based virtual screening of chemical databases. II. Are homology models of G-protein-coupled receptors suitable targets? *Proteins* 50, 5–25

57 Schafferhans, A. *et al.* (2001) Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* 307, 407–427

58 Oshiro, C. *et al.* (2004) Performance of 3D database molecular docking studies into homology models. *J. Med. Chem.* 47, 764–767

59 Steinrücke, P. *et al.* (2000) Design of helical proteins for real-time endoprotease assays. *Anal. Biochem.* 286, 26–34

60 Benhamou, B. *et al.* (1992) A single amino acid that determines the sensitivity of progesterone receptors to RU486. *Science* 255, 206–209

61 Gray, G.O. *et al.* (1987) RU486 is not an antiprogestin in the hamster. *J. Steroid Biochem.* 28, 493–497

62 Williams, P.A. *et al.* (2000) Mammalian microsomal cytochrome P450 monooxygenase: structural adaptations for membrane binding and functional diversity. *Mol. Cell* 5, 121–131

63 Schoch, G.A. *et al.* (2004) Structure of human microsomal cytochrome P450 2C8. Evidence for a peripheral fatty acid binding site. *J. Biol. Chem.* 279, 9497–9503

64 Williams, P.A. *et al.* (2003) Crystal structure of human cytochrome P450 2C9 with bound warfarin. *Nature* 424, 464–468

65 Lewis, D.F. (2002) Molecular modeling of human cytochrome P450-substrate interactions. *Drug Metab. Rev.* 34, 55–67

66 Lewis, D.F. (2002) Modelling human cytochromes P450 involved in drug metabolism from the CYP2C5 crystallographic template. *J. Inorg. Biochem.* 91, 502–514

67 Lewis, D.F. *et al.* (2003) Homology modelling of human CYP1A2 based on the CYP2C5 crystallographic template structure. *Xenobiotica* 33, 239–254

68 Lewis, D.F. *et al.* (2003) Homology modelling of CYP2A6 based on the CYP2C5 crystallographic template: enzyme–substrate interactions and QSARs for binding affinity and inhibition. *Toxicol. In Vitro* 17, 179–190

69 Lewis, D.F. *et al.* (2002) Molecular modelling of CYP2B6 based on homology with the CYP2C5 crystal structure: analysis of enzyme–substrate interactions. *Drug Metabol. Drug Interact.* 19, 115–135

70 Lewis, D.F. *et al.* (2003) A molecular model of CYP2D6 constructed by homology with the CYP2C5 crystallographic template: investigation of enzyme–substrate interactions. *Drug Metabol. Drug Interact.* 19, 189–210

71 Lewis, D.F. *et al.* (2003) Investigation of enzyme selectivity in the human CYP2C subfamily: homology modelling of CYP2C8, CYP2C9 and CYP2C19 from the CYP2C5 crystallographic template. *Drug Metabol. Drug Interact.* 19, 257–285

72 Vermeulen, N.P. (2003) Prediction of drug metabolism: the case of cytochrome P450 2D6. *Curr. Top. Med. Chem.* 3, 1227–1239

73 De Groot, M.J. *et al.* (1999) A novel approach to predicting P450 mediated drug metabolism. CYP2D6 catalyzed *N*-dealkylation reactions and qualitative metabolite predictions using a combined protein and pharmacophore model for CYP2D6. *J. Med. Chem.* 42, 4062–4070

74 Afzelius, L. *et al.* (2001) Competitive CYP2C9 inhibitors: enzyme inhibition studies, protein homology modeling, and three-dimensional quantitative structure–activity relationship analysis. *Mol. Pharmacol.* 59, 909–919

75 Szklarz, G.D. *et al.* (1998) Molecular basis of P450 inhibition and activation: implications for drug development and drug therapy. *Drug Metab. Dispos.* 26, 1179–1184

76 Mankowskia, D.C. *et al.* (2003) Prediction of human drug metabolizing enzyme induction. *Curr. Drug Metab.* 4, 381–391

77 Watkins, R.E. *et al.* (2003) Coactivator binding promotes the specific interaction between ligand and the pregnane X receptor. *J. Mol. Biol.* 331, 815–828

78 Dussault, I. *et al.* (2002) A structural model of the constitutive androstane receptor defines novel interactions that mediate ligand-independent activity. *Mol. Cell. Biol.* 22, 5270–5280

79 Xiao, L. *et al.* (2002) Insights from a three-dimensional model into ligand binding to constitutive active receptor. *Drug Metab. Dispos.* 30, 951–956

80 Lewis, D.F. *et al.* (2002) Molecular modelling of the human glucocorticoid receptor (hGR) ligand-binding domain (LBD) by homology with the human estrogen receptor alpha (hERalpha) LBD: quantitative structure-activity relationships within a series of CYP3A4 inducers where induction is mediated via hGR involvement. *J. Steroid Biochem. Mol. Biol.* 82, 195–199

81 Ekins, S. (2004) Predicting undesirable drug interactions with promiscuous proteins *in silico*. *Drug Discov. Today* 9, 276–285

82 Mitcheson, J.S. *et al.* (2000) A structural basis for drug-induced long QT syndrome. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12329–12333

83 Pearlstein, R.A. *et al.* (2003) Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg. Med. Chem. Lett.* 13, 1829–1835

84 Hillisch, A. and Hilgenfeld, R. (2003) *Modern Methods of Drug Discovery,* Birkhäuser Verlag AG