Homology modeling in drug discovery: current trends and applications

Claudio N. Cavasotto and Sharangdhar S. Phatak

School of Health Information Sciences, The University of Texas Health Science Center at Houston, 7000 Fannin, Suite 860B, Houston, TX 77030, United States

As structural genomics (SG) projects continue to deposit representative 3D structures of proteins, homology modeling methods will play an increasing role in structure-based drug discovery. Although computational structure prediction methods provide a cost-effective alternative in the absence of experimental structures, developing accurate enough models still remains a big challenge. In this contribution, we report the current developments in this field, discuss *in silico* modeling limitations, and review the successful application of this technique to different stages of the drug discovery process.

Historically, the drug discovery process has relied on experimental high-throughput screening (HTS) to identify biologically active compounds [1]. Despite advances in automation techniques for HTS this approach remains extremely laborious, expensive and has often failed to identify potent lead series [2,3]. Complementary *in silico* methods like structure-based drug design (SBDD), incorporate the knowledge from high-resolution 3D protein structures to probe structure–function relationships [4], identify and select therapeutically relevant targets (assess druggability) [4,5], study the molecular basis of ligand:protein interactions [5], characterize binding pockets [6], develop target-specific compound libraries [7,8], identify hits by high-throughput docking (HTD) [9,10] and optimize lead compounds [11], all of which can be used to rationalize, increase efficiency, speed and cost-effectiveness of the drug discovery process [1,12,13].

Structural information of biological macromolecules is readily available in the Protein Data Bank (PDB), http://www.pdb.org. By April 2009, the PDB contained ~57,000 experimental protein structures, that can be grouped into ~3500 families, consisting of nearly 1100 unique folds (http://scop.mrc-lmb.cam.ac.uk/scop/count.html#scop-1.73). Considering that the number of nonredundant amino acid sequence entries is around 408,000 (http://www.expasy.org/sprot/), this shows the huge gap between known annotated sequences and available 3D structures. Despite the rapid growth of the PDB, the structural novelty of proteins (defined as <25% sequence identity between 2 structures) deposited in the PDB has remained constant since 1992 [14]. To have at least one representative structure from protein families with no experimental structural information, several structural genomics (SG) projects were initiated, which have contributed roughly 50% of the novel structures (defined as <30% sequence identity) in the PDB over the past five years [11,14].

SG projects have also spurred the developments in X-ray crystallography and NMR techniques, so that new protein structures are solved quicker, which in turn has widened their use in drug discovery [12,15–17]. Still, the bottleneck for the determination of the 3D protein structure is problems associated with the purification and crystallization of proteins, for example G-protein-coupled receptors (GPCRs), among others.

In the absence of experimental structures, computational methods are used to predict 3D protein models to provide insight into the structure and function of these proteins. Repositories like The SWISS-MODEL (http://swissmodel.expasy.org/SWISS-MODEL.html), Protein Model Portal (http://proteinmodelportal.org) [18] and Modbase (http://modbase.compbio.ucsf.edu) [19], contain protein models generated using various automated methods. These servers provide models that serve as starting points for biologists/experimentalists to assist in SG and biomedical research projects. However, without human intervention, errors as a result of inaccurate sequence alignment, and inability to identify and correctly model domains, such as loop and ligand-binding regions, are magnified, which results in low-accuracy generated models, thus limiting their applicability to drug discovery projects [20,21]. The development and improvement of homology modeling

Corresponding author: Cavasotto, C.N. (Claudio.N.Cavasotto@uth.tmc.edu)

1359-6446/06/$ - see front matter © 2009 Elsevier Ltd. All rights reserved. doi:10.1016/j.drudis.2009.04.006

refinement tools in the context of drug discovery is an active area of research [22]. In this review, we highlight recently developed homology modeling methods and the applications of this methodology in the drug discovery process.

Homology or comparative modeling

Comparative or homology modeling is a methodology to predict protein structure based on the general observation that proteins with similar sequences have similar structures. Given an experimentally established protein structure (template), models can be generated for a homologous sequence (target) that shares with either the template significant sequence (~30% or more) or structural similarity (e.g. Class A GPCRs share a common seven-transmembrane helical structure, despite low sequence similarities between family members). The process of homology or comparative modeling of proteins consists of the following steps (Fig. 1): (1) identification of known 3D structure(s) of a related protein that can serve as template; (2) sequence alignment of target and template proteins; (3) model building for the target based on the 3D structure of the template and the alignment; (4) refining/validation/evaluation of the models. These steps may be repeated until a satisfactory model is built [23].

Models, by definition, are an abstraction and hence may contain errors. Depending on the degree of sequence identity or similarity, and the quality of the alignment, the accuracy of homology models compared to the actual experimental structure can be up to ~1–2 Å C_α atom RMSD (root-mean-square deviation distance between corresponding C_α atoms) [24,25]. As a general rule, models built with over 50% sequence identities are accurate enough for drug discovery applications, those between 25 and 50% identities can be used to assess target druggability and design mutagenesis experiments, and those in between 10 and 25% are speculative at best [4]. Although model quality is directly related to the identity (or similarity) between template and target sequences, this rule does not always hold. In GPCRs, for example, low sequence identities among family members are compensated by a common seven-transmembrane helical fold, a structural similarity which can be exploited for modeling purposes. Conversely, an overall high sequence identity might mask dissimilarities in certain regions like exposed loops, which are most likely to be flexible, adding uncertainty to the model, and thereby rendering it of lesser value for drug discovery applications. The choice of template, inaccurate alignments and inefficient refinement methods are the main sources of errors in homology modeling [26].

Sequence alignment plays an important role in developing accurate homology models. Pairwise alignment tools implement dynamic programming methods to search for optimum alignments (local or global) between a pair of sequences. This approach is useful to search databases for homologous sequences. Multiple sequence alignment methods simultaneously align several sequences to identify conserved regions, predict functional sites and protein function as well as aid phylogenetic analysis [27]. This approach is particularly suited for proteins with low (<40%) sequence identities. In cases where a single template fails to provide complete structural information of the target, the multi-template approach [28], or mixing and matching techniques [29]

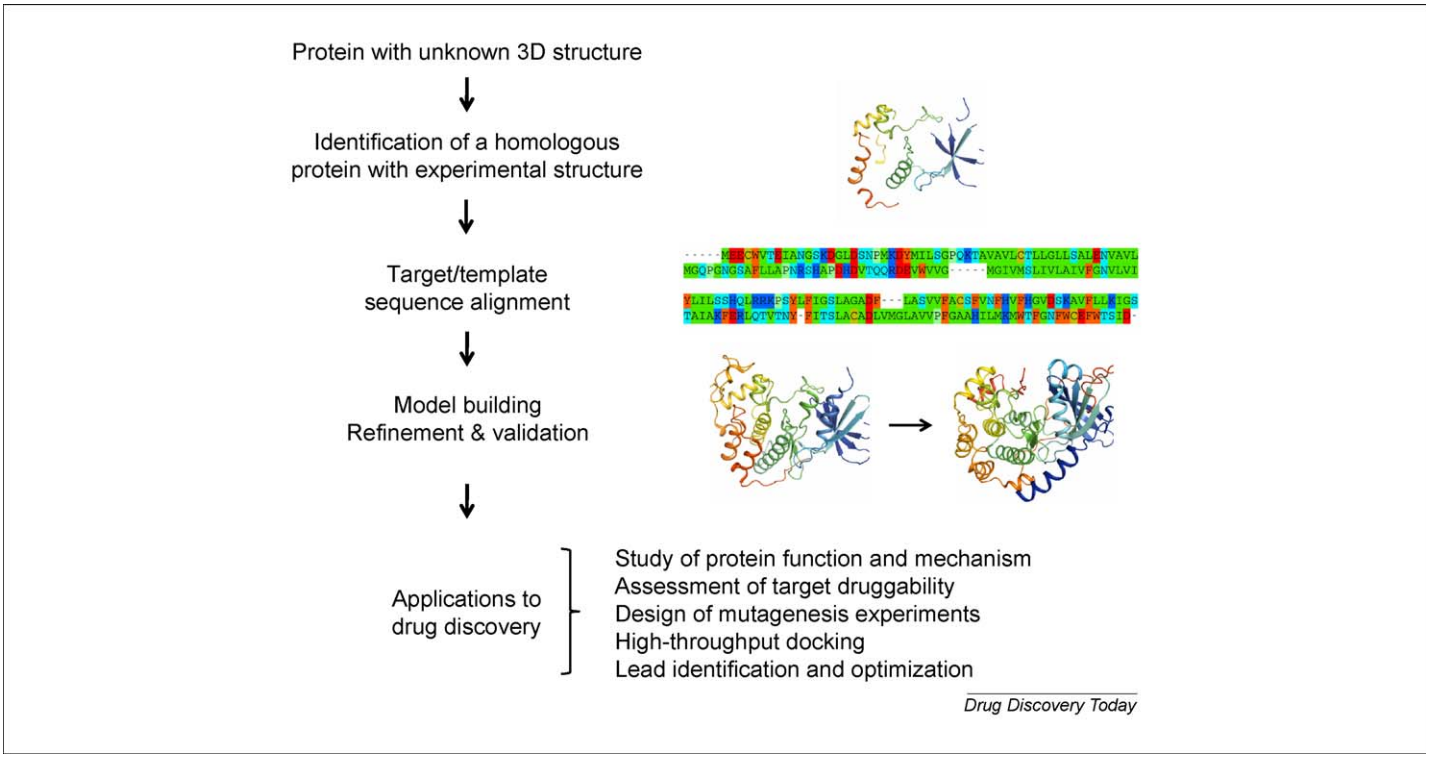

FIGURE 1

Outline of the homology modeling process and its applications in drug discovery. Given the sequence of a protein with unknown structure, the first step is the identification of a related protein with known 3D structure that serves as template. An alignment of the target and template sequences is necessary to assign the correspondence between target and template residues. A model is then built for the target based on the alignment and structure of the template, and further refined and validated. This figure was prepared using Pymol (http://www.pymol.org).

TABLE 1

Applications of homology modeling techniques at different stages of the drug discovery process within the past two years

Receptor	Application	Refs
BC0371 (member of the enolase superfamily)	Study of protein function	[35]
M antigen	Study of protein function	[37]
RDH12	Assess the biological role of mutations in the binding site and their effect on the function of RDH12	[38]
Nod-like receptors	Understand the protein mechanism implicated in the immune response	[39]
Glut 1 transport receptor	Provide insights into the transport mechanism	[40]
NHE1	Understand the access mechanism of Na^+/H^+ exchange	[41]
UreF	Study of protein mechanism	[42]
Prothrombinase (FXa–FVA)	Understand prothrombinase function and identify druggable pockets	[43]
ECE-2	Understand the loss of catalytic activity of ECE-2 via mutagenesis studies	[48]
Lunatic fringe (Lfng)	Explain the binding mode of *O*-linked fucose	[49]
MCH-R1	Structure-based lead discovery for antiobesity drugs	[52]
RSK-2	Docking-based screening for breast and prostate cancer lead discovery	[59]
Cannabinoid receptor 2, human adenosine A2A receptor, alpha-1-A-adrenoreceptor, adenosine 3 receptor	Binding mode prediction and elucidation of ligand:protein interactions	[60–63]
Peptide CGP38560 in complex with renin, Src kinase, PKC theta, GPR109A	Lead or compound optimization	[64,66–68]
Alpha-glucosidase	Structure-based virtual screening for antidiabetes lead discovery	[69]
Cdc25 phosphatases	Structure-based virtual screening for anticancer lead discovery	[70]
Protein tyrosine phosphatase SHP2	Structure-based virtual screening for anticancer lead discovery	[71]
CK1δ	Structure-based virtual screening to identify inhibitors for Alzheimer's disease	[72]
6-Phosphofructo-2-kinase (PFKFB3)	Structure-based virtual screening for inhibitors against glycolytic flux and tumor growth.	[73]

incorporate structural information from multiple homologous templates to improve the overall model quality. A successful example of a multitemplate approach is the modeling of the breast cancer resistance protein (ABCG2), in which two different domains of the protein (NB and TM) were modeled using two different templates, to study receptor–ligand interactions [30].

Use of homology models to characterize protein structural properties, function and mechanisms

Structure-based tools to explain protein function and interaction patterns

A clear understanding of the function and physiological role of a therapeutic target is a prerequisite for a successful drug discovery program. A potential target with a known physiological function may not serve as a successful target for drug discovery simply because its mechanism of action may not be suited for small molecule targeting, for example, some protein:protein interactions [31]. The process of understanding both the function and mechanism of action, termed target identification and validation, is of utmost importance before commencing any drug discovery program.

Functional annotation of protein sequences based on sequence similarities is an ambiguous process as there is no direct relationship between sequence similarities and function [32]. With the increase in availability of 3D protein structures, several methods have been developed, which supplement sequence information with structural information to probe biological function [33,34]. In the absence of experimentally qualified protein structures, there are several examples where homology models have aided in the prediction of protein function.

Song *et al.* combined experimental methods and homology modeling to predict the function of BC0371, a member of the functionally diverse enolase superfamily [35] (recent applications of homology modeling relevant to the drug discovery process are listed in Table 1). Phylogenetic analysis associated BC0371 with the subgroup of the L-Ala-D/L-Glu epimerase (AEE), whose physiological function is the interconversion of L-Ala-D-Glu into L-Ala-L-Glu. However, the homology model of BC0371 constructed using the crystal structure of AEE from *Bacillus subtilis*, revealed key differences in the binding pocket of BC0371 and AEE, thus casting doubt into the results from the phylogenetic analysis. In a previous work, a docking and scoring protocol was specifically developed for the enolase family, which was able to correctly rank the known substrates of functionally assigned and structurally characterized enolase family members, thus showing the ability to capture substrate binding selectivity [36]. On the basis of the above hypotheses, *in silico* docking using metabolite ligand libraries revealed that BC0371 ranked the *N*-succinyl-L-amino acid substrates significantly higher than other ligands, which was consistent with the experimental results that predicted the function of BC0371 to be the racemization of *N*-succinyl-L-arginine/lysine (NSAR), thus implying that BC0371 is not part of the AEE subgroup family, but of the NSAR subgroup. In another study, Nosanchuk *et al.* developed a homology model for the M antigen, a major

TABLE 2

Recent developments in homology modeling methods relevant to drug discovery

Method	Description	Refs
Homology modeling-based complex prediction (HOMBACOP)	Prediction of protein:protein complexes to understand mechanisms and interactions patterns	[45]
Propensity for ligand binding index (PLB)	Identification of druggable binding sites in homology models	[47]
Ligand-steered homology modeling of the binding site	To characterize binding sites by using known ligands and experimental constraints, through a flexible-ligand:flexible-receptor docking-based stochastic global energy minimization in the internal coordinate space	[52]
Ligand-based homology modeling with binary scoring function	To reproduce ligand-binding modes through manual optimization of side chains based on experimental studies, and use the ligand:receptor interaction pattern to develop a binary scoring function for virtual screening applications	[57]
Molecular dynamics based methods	To optimize binding sites via pressure-based steered dynamics or constrained molecular dynamics	[58,59]

diagnostic antigen of *Histoplasma capsulatum* (HC), to predict its function in the pathogenesis of histoplasmosis, caused by HC. Although the M antigen model closely resembled enzymes that evade oxidative stress, modeling studies indicated that the M antigen is localized on the cell surface of HC and sterically accessible to antibodies, suggesting its role in the diagnosis of histoplasmosis [37]. A homology model of the RDH12 gene, implicated in Leber congenital amaurosis (LCA), a leading cause of inherited childhood blindness, was instrumental in explaining the role of mutations in the binding site and subsequent inactivation of RDH12 [38].

Homology models have also been used to aid in the understanding of protein mechanisms. The Nod-like receptor (NLR) family is implicated in the innate immune response. Homology modeling and subsequent experimental analyses showed that the NLR family has a domain structure similar to the apoptotic inhibitor protein Apaf-1, which forms a heptameric platform to trigger apoptosis, thus suggesting a similar biochemical behavior for NLR [39]. Homology modeling and mutagenesis studies have provided insights into the role in the transport mechanism of transmembrane helix 6 of the Glut 1 glucose transporter. The studies indicate that residues in transmembrane helix 6 do not interact with the solvent, so helix 6 is likely to be associated with the transport cycle and not with substrate binding [40]. On the basis of the *in silico* models of the human Na^+/H^+ exchanger 1 (NHE1), Ben-Tal *et al.* proposed an alternative access mechanism of Na^+/ H^+ exchange in NHE1 [41], and 3D modeling techniques helped to understand the mechanism of UreF protein as a GTPase activating protein in the urease active site biosynthesis [42]. To understand the prothrombinase function, and to assess potential druggable pockets at the protein interfaces for therapeutic interventions, Villoutreix *et al.* developed structural models of prothrombinase (FXa–FVa) explaining the role of several residues relevant to the FXa–FVa interactions [43].

Protein–protein interactions (PPIs) are crucial for many disease-related cellular processes and these complexes represent one of the biggest challenges in drug discovery [31]. Taking advantage of the increasing number of experimental structures for such complexes, Alexov *et al.* have developed a homology modeling based method, HOMBACOP (homology-based complex prediction), to predict 3D protein:protein complexes and understand the mechanisms and interactions at the cellular level, for chemical biology or drug discovery projects [44,45]. Refer to Table 2 for a list of recent methodological developments in homology modeling in the context of drug discovery.

Assessing target druggability from in silico *generated structures*

A druggable target has the ability to tightly bind with small molecules. As most drugs bind to specific binding sites on a protein, it makes sense to identify *a priori* such domains as a measure of target druggability. Researchers at Eidogen-Sertanty developed the Target Informatics Platform (TIP), which contains information about protein structure/homology models and binding sites of several protein families [46]. It is important to note that complementing crystal structures with homology models has resulted in 100% structural coverage of some gene families like Nuclear Receptors, Phosphodiesterase and over 98% coverage of protein kinases and trypsin-like proteases [46]. Using TIP and the complex of COX-2 with its inhibitor celecoxib, researchers were able to identify a similar binding site in the PPARγ receptor which contained several important binding residues, offering possible clues to design novel PPAR ligands. In another study, Hirayama *et al.* developed an index termed propensity for ligand binding (PLB) to identify druggable binding sites in homology models, which was later used to successfully predict the druggable cavity in a homology model of tryptophanyl-tRNA synthetase [47].

Design of mutagenesis experiments using three-dimensional structures

Mutagenesis studies play an important role in the identification of amino acids with relevant biological function. The structural information obtained from homology modeling helps in rationalizing the selection of amino acids for such studies. There are several reports where homology models have aided mutagenesis experiments to study ligand:receptor interactions, analyze the role of nonconserved residues in active sites and suggest ligand binding modes. A homology model of the endothelin-converting enzyme-2 (ECE-2), a neuropeptide-processing enzyme, was built based on the template of the homologous neprilysin receptor to understand the loss of catalytic activity and substrate/inhibitor binding of ECE-2. On the basis of the structural insights obtained from the model, mutagenesis experiments on the two nonconserved residues in the binding pocket of ECE-2 (Y563 and W148) suggested that Y563 may be responsible for the loss of catalytic activity of ECE-2 for [48]. Haliwanger *et al.* used the structural knowledge

obtained from the homology model of lunatic fringe (Lfng) to conduct mutagenesis experiments on the active site residues to explain the binding mode of *O*-linked fucose [49].

Homology modeling: methods and application in lead identification

Cheaper, faster computational resources, increasing availability of high resolution protein structures or *in silico* models, and advancement in docking algorithms have made HTD a tool of choice to filter out promising compounds with relevant chemical functionalities from large virtual chemical libraries for further experimental evaluation. Clearly, in the absence of crystal structures, homology models are the only alternative to get a 3D representation of the target. Although homology-modeling methods can build reasonably accurate models, refinement methods are needed to get a more accurate characterization of the binding site, and determine the exact side chain conformation, as minor errors may render the model useless for HTD applications.

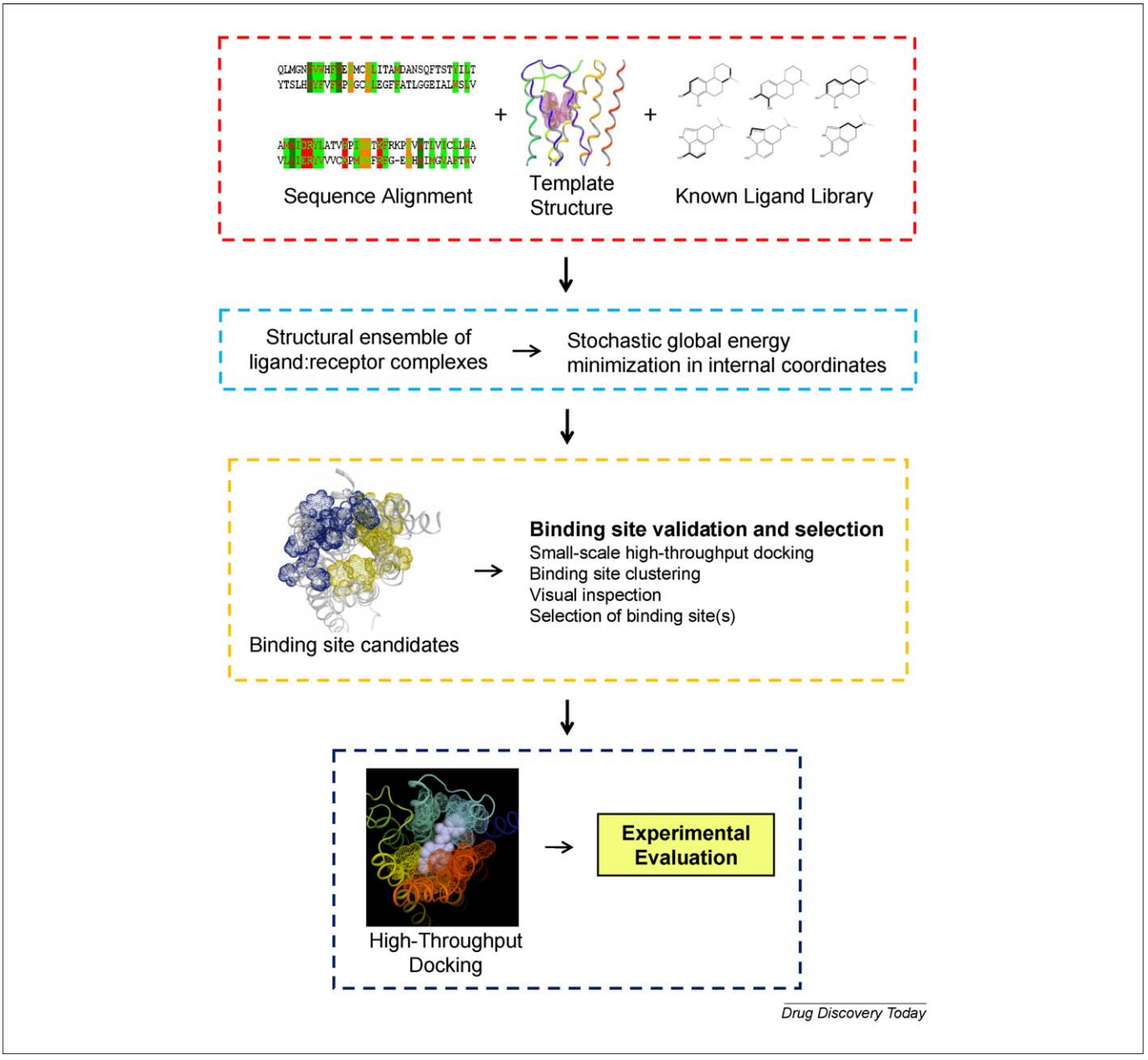

FIGURE 2

Outline of the ligand-steered homology modeling method [52], in which information of known ligands is explicitly used to shape and optimize the binding site. Starting with the alignment between the template and target sequences, the structure of the target and a collection of known ligands of the target (red border), a structural ensemble of ligand:receptor complexes is generated and optimized through a stochastic global energy minimization in internal coordinates, eventually including experimental constraints (light blue border). Binding site candidates thus generated are validated through binding site clustering, small-scale high-throughput docking, analysis of known experimental interactions (if any), and visual inspection; one or more binding sites are selected (orange border), which could be later used in structure-based drug discovery projects (dark blue border). In Ref. [52] models of the melanin-concentrating hormone receptor 1 (MCHR1) were generated with this method and used for high-throughput docking followed by experimental evaluation of top-ranking hits. Six low micromolar antagonists were found. This figure was prepared using Pymol (http://www.pymol.org) and Maestro (Version 8.5, Schrödinger, LLC, New York, NY, 2008).

GPCRs constitute the largest family of signaling receptors in the cell and are the target for nearly half of all drug discovery programs [50]. Until recently, there was only one crystal structure available, bovine rhodopsin (bRho) (PDB code 1f88, 1l9h), what hampered a wider structure-based drug design approach for GPCRs. Recently, the appearance of crystal structures for four new GPCRs (Opsin, 3cap; β2 adrenergic (β2-AR), 2rh1; turkey β1 adrenergic (β1-AR), 2vt4; human A2A adenosine receptor, 3eml) brings a broader template diversity for *in silico* modeling. The newly crystallized β2-AR has been already investigated as an alternative template to model other Class-A GPCRs for drug discovery applications [51].

Recently, a novel ligand-steered homology modeling method was presented [52] (Fig. 2), in which information about known ligands is explicitly used to shape and optimize the binding site through a docking-based stochastic global energy minimization procedure in the internal coordinate space [53–56]. This method is particularly useful to reduce the uncertainty in modeling the binding site, as both the ligand and receptor are held flexible during the modeling stage. The ligand-steered modeling method was developed and tested on the melanin-concentrating hormone receptor 1 (MCH-R1), a target for antiobesity drugs. Models thus generated were validated through a small-scale HTD, and used then in a large-scale HTD, followed by competition assays of the top-ranking hits. Thus, six novel low-micromolar hits were found. It is emphasized that the hit-enrichment rate obtained in this project was tenfold better as compared to the traditional HTS hit-rate [52].

In another ligand-based homology modeling approach for GPCRs, the side chains within the binding site of the mGluR5 were manually adjusted in the presence of the known inhibitors. The ligand:receptor interactions from these complexes were used to develop an unbiased binary fingerprint scoring function (IFS), which fared better as compared to several conventional scoring functions in ranking antagonists of mGluR5 obtained from virtual screening experiments [57].

Another binding site optimization approach for GPCRs by Kimura *et al.* shapes and optimizes the binding pockets for chemokine receptor-2 (CCR2) by using a pressure-based steered molecular dynamics method. Small radii Lennard–Jones particles are placed on a grid and tethered via weak harmonic bonds to four nearest neighbors within the binding pocket. The backbone dihedral angles of the helices are restrained, and the radii of the Lennard–Jones particles are increased during MD simulation to mimic increased pressure and expand the binding pocket. Models were validated by their ability to redock three known antagonists for the CCR2 receptor. This method does not include any ligand information and can potentially be applied in lead identification studies for several receptor types [58].

A combination of homology modeling, constrained molecular dynamics and known inhibitors crystallized with other homologous proteins was used to shape and optimize the binding site of the ribosomal S6 kinase 2 (RSK2), a potential drug target for human breast and prostate cancer. Two low micromolar inhibitors were obtained from subsequent docking-based screening study [59].

Homology-based lead optimization and SAR rationalization

Lead optimization is a lengthy iterative process of modifying the chemical structure of a known hit to modify its physico-chemical and pharmacological properties to improve bioavailability, reduce unwanted toxicity and achieve the required drug profile suitable for animal model studies and clinical trials. Understanding ligand:protein interaction is a key step for lead optimization in SBDD projects, because the structural insights obtained can guide the optimization of ligands toward improved pharmacological profiles.

Homology models have proved to be an invaluable source to rationalize SAR data and predict binding modes of experimental compounds for Cannabinoid receptor 2 [60,74], human adenosine A2A receptor [61] and alpha-1-adrenoreceptors [62]. Moro *et al.* successfully explained the activities of the pyrazolotriazolopyrimidines class of inhibitors for the human adenosine 3 (A3) receptors by refining the binding site of A3 [63]. Researchers at Ciba-Geigy (now Novartis) used a homology model complex of the peptide inhibitor (CGP38560) with renin to identify the bioactive conformation of the peptide and putative hydrogen bonds with the receptor. This information was highly useful to design and launch the first renin inhibitor drug Aliskiren [64]. Using the crystal structure of rat farnesyl protein transferase, Schlitzer *et al.* generated a homology model of *Plasmodium falciparum* to study active site differences in mammalian and parasitic enzymes, and to optimize the farnesyl transferase class of antimalarial inhibitors into more selective benzophenone-based inhibitors [65]. Noronha *et al.* developed a homology model for Src-kinase to analyze the binding mode of benzotriazine compounds and incorporated that knowledge to design and optimize a 1 μM inhibitor with higher selectivity and better pharmacokinetic properties as compared to the other known benzotrizine-based compounds [66]. Homology modeling and structure–activity relationship (SAR) data were instrumental in the development of potent and selective compounds for protein kinase C theta (PKCθ), an important target for the treatment of psoriasis [67]. Homology models have also aided lead optimization in the nicotinic receptor GPR109A agonists [68]. A complete list of recent applications of homology modeling to drug discovery is displayed in Table 1.

Conclusions

In the absence of experimental structures, homology modeling plays an important role in the structure-based drug discovery process. Despite clear achievements of the SG projects, and the rapid technological advances in crystallization and NMR methods, it is reasonable to assume that experimental structures for most of the therapeutically relevant targets will not be available in the near future. Thus, *in silico* homology modeling provides a viable cost-effective alternative to generate reasonably accurate models for drug discovery. To date, homology modeling has been successfully used to identify hits using HTD, to suggest accurate binding modes and ligand:receptor interactions, aid in mutagenesis experiments, rationalize SAR data and guide the optimization toward more potent ligands. However problems associated with template identification, accurate sequence alignment and refinement methods hinder a wider use of *in silico* generated models in the drug discovery process. Thus, there is a need to develop better methods to reduce modeling errors, specifically at pharmacologically relevant sites, and when target and template proteins share lower sequence

similarities. In this review, we have summarized current advances in homology modeling methods, and reported the latest applications of this technique to the drug discovery process.

References

1 Congreve, M. *et al.* (2005) Structural biology and drug discovery. *Drug Discov. Today* 10, 895–907
2 Kubinyi, H. (2001) HTS technologies – IBC informa conference. *IDrugs* 4, 168–173
3 Liu, B. *et al.* (2004) Technological advances in high-throughput screening. *Am. J. Pharmacogenom.* 4, 263–276
4 Hillisch, A. *et al.* (2004) Utility of homology models in the drug discovery process. *Drug Discov. Today* 9, 659–669
5 Blundell, T.L. *et al.* (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos. Trans. R. Soc. Lond. B: Biol. Sci.* 361, 413–423
6 Halgren, T.A. (2009) Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* 49, 337–389
7 Orry, A.J. *et al.* (2006) Structure-based development of target-specific compound libraries. *Drug Discov. Today* 11, 261–266
8 Schnur, D.M. (2008) Recent trends in library design: 'rational design' revisited. *Curr. Opin. Drug Discov. Dev.* 11, 375–380
9 Cavasotto, C.N. and Orry, A.J. (2007) Ligand docking and structure-based virtual screening in drug discovery. *Curr. Top. Med. Chem.* 7, 1006–1014
10 Cavasotto, C.N. and Singh, N. (2008) Docking and high throughput docking: successes and the challenge of protein flexibility. *Curr. Comput. Aid. Drug Discov.* 4, 221–234
11 Lundstrom, K. (2007) Structural genomics and drug discovery. *J. Cell Mol. Med.* 11, 224–238
12 Anderson, A.C. (2003) The process of structure-based drug design. *Chem. Biol.* 10, 787–797
13 Weigelt, J. *et al.* (2008) Structural genomics and drug discovery: all in the family. *Curr. Opin. Chem. Biol.* 12, 32–39
14 Levitt, M. (2007) Growth of novel protein structural data. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3183–3188
15 Manjasetty, B.A. *et al.* (2008) Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics* 8, 612–625
16 Mooij, W.T. *et al.* (2006) Automated protein-ligand crystallography for structure-based drug design. *ChemMedChem* 1, 827–838
17 Villar, H.O. *et al.* (2004) Using NMR for ligand discovery and optimization. *Curr. Opin. Chem. Biol.* 8, 387–391
18 Arnold, K. *et al.* (2009) The protein model portal. *J. Struct. Funct. Genom.* 10, 1–8
19 Pieper, U. *et al.* (2006) MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* 34 (Database issue), D291–295
20 Dalton, J.A. and Jackson, R.M. (2007) An evaluation of automated homology modelling methods at low target template sequence similarity. *Bioinformatics* 23, 1901–1908
21 Fischer, D. (2006) Servers for protein structure prediction. *Curr. Opin. Struct. Biol.* 16, 178–182
22 Kairys, V. *et al.* (2006) Using protein homology models for structure-based studies: approaches to model refinement. *Sci. World J.* 6, 1542–1554
23 Marti-Renom, M.A. *et al.* (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291–325
24 Burley, S.K. *et al.* (2008) Contributions to the NIH-NIGMS protein structure initiative from the PSI production centers. *Structure* 16, 5–11
25 Ginalski, K. (2006) Comparative modeling for protein structure prediction. *Curr. Opin. Struct. Biol.* 16, 172–177
26 Larsson, P. *et al.* (2008) Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Sci.* 17, 990–1002
27 Edgar, R.C. and Batzoglou, S. (2006) Multiple sequence alignment. *Curr. Opin. Struct. Biol.* 16, 368–373
28 Cheng, J. (2008) A multi-template combination algorithm for protein comparative modeling. *BMC Struct. Biol.* 8, 18
29 Liu, T. *et al.* (2008) Improving the accuracy of template-based predictions by mixing and matching between initial models. *BMC Struct. Biol.* 8, 24
30 Hazai, E. and Bikadi, Z. (2008) Homology modeling of breast cancer resistance protein (ABCG2). *J. Struct. Biol.* 162, 63–74
31 Fry, D.C. (2006) Protein–protein interactions as targets for small molecule drug discovery. *Biopolymers* 84, 535–552
32 Tian, W. and Skolnick, J. (2003) How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol.* 333, 863–882
33 Cammer, S.A. *et al.* (2003) Structure-based active site profiles for genome analysis and functional family subclassification. *J. Mol. Biol.* 334, 387–401
34 Wangikar, P.P. *et al.* (2003) Functional sites in protein families uncovered via an objective and automated graph theoretic approach. *J. Mol. Biol.* 326, 955–978
35 Song, L. *et al.* (2007) Prediction and assignment of function for a divergent *N*-succinyl amino acid racemase. *Nat. Chem. Biol.* 3, 486–491
36 Kalyanaraman, C. *et al.* (2005) Virtual screening against highly charged active sites: identifying substrates of alpha–beta barrel enzymes. *Biochemistry* 44, 2059–2071
37 Guimaraes, A.J. *et al.* (2008) Biological function and molecular mapping of M antigen in yeast phase of *Histoplasma capsulatum*. *PLoS ONE* 3, e3449
38 Sun, W. *et al.* (2007) Novel RDH12 mutations associated with Leber congenital amaurosis and cone–rod dystrophy: biochemical and clinical evaluations. *Vision Res.* 47, 2055–2066
39 Proell, M. *et al.* (2008) The Nod-like receptor (NLR) family: a tale of similarities and differences. *PLoS ONE* 3, e2119
40 Mueckler, M. and Makepeace, C. (2008) Transmembrane segment 6 of the Glut1 glucose transporter is an outer helix and contains amino acid side chains essential for transport activity. *J. Biol. Chem.* 283, 11550–11555
41 Landau, M. *et al.* (2007) Model structure of the Na^+/H^+ exchanger 1 (NHE1): functional and clinical implications. *J. Biol. Chem.* 282, 37854–37863
42 Salomone-Stagni, M. *et al.* (2007) A model-based proposal for the role of UreF as a GTPase-activating protein in the urease active site biosynthesis. *Proteins* 68, 749–761
43 Autin, L. *et al.* (2006) Proposed structural models of the prothrombinase (FXa–FVa) complex. *Proteins* 63, 440–450
44 Kundrotas, P.J. and Alexov, E. (2006) Predicting 3D structures of transient protein–protein complexes by homology. *Biochim. Biophys. Acta* 1764, 1498–1511
45 Kundrotas, P.J. *et al.* (2008) Homology-based modeling of 3D structures of protein–protein complexes using alignments of modified sequence profiles. *Int. J. Biol. Macromol.* 43, 198–208
46 Hambly, K. *et al.* (2006) Interrogating the druggable genome with structural informatics. *Mol. Divers.* 10, 273–281
47 Soga, S. *et al.* (2007) Identification of the druggable concavity in homology models using the PLB index. *J. Chem. Inf. Model.* 47, 2287–2292
48 Gagnidze, K. *et al.* (2008) Homology modeling and site-directed mutagenesis to identify selective inhibitors of endothelin-converting enzyme-2. *J. Med. Chem.* 51, 3378–3387
49 Luther, K.B. *et al.* (2009) Structural and mechanistic insights into lunatic fringe from a kinetic analysis of enzyme mutants. *J. Biol. Chem.* 284, 3294–3305
50 Wolf, S. *et al.* (2008) Simulations of a G protein-coupled receptor homology model predict dynamic features and a ligand binding site. *FEBS Lett.* 582, 3335–3342
51 Costanzi, S. (2008) On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor. *J. Med. Chem.* 51, 2907–2914
52 Cavasotto, C.N. *et al.* (2008) Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* 51, 581–588
53 Cavasotto, C.N. and Abagyan, R.A. (2004) Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* 337, 209–225
54 Kovacs, J.A. *et al.* (2005) Conformational sampling of protein flexibility in generalized coordinates: application to ligand docking. *J. Comput. Theor. Nanosci.* 2, 354–361
55 Monti, M.C. *et al.* (2007) Scalaradial, a dialdehyde-containing marine metabolite that causes an unexpected noncovalent PLA2 inactivation. *ChemBioChem* 8, 1585–1591
56 Monti, M.C. *et al.* (2009) The binding mode of petrosaspongiolide M to the human group IIA phospholipase A(2): exploring the role of covalent and noncovalent interactions in the inhibition process. *Chemistry* 15, 1155–1163
57 Radestock, S. *et al.* (2008) Homology model-based virtual screening for GPCR ligands using docking and target-biased scoring. *J. Chem. Inf. Model.* 48, 1104–1117
58 Kimura, S.R. *et al.* (2008) Expanding GPCR homology model binding sites via a balloon potential: a molecular dynamics refinement approach. *Proteins* 71, 1919–1929
59 Nguyen, T.L. *et al.* (2006) Homology model of RSK2 N-terminal kinase domain, structure-based identification of novel RSK2 inhibitors, and preliminary common pharmacophore. *Bioorg. Med. Chem.* 14, 6097–6105

60 Diaz, P. *et al.* (2009) 6-Methoxy-*N*-alkyl isatin acylhydrazone derivatives as a novel series of potent selective cannabinoid receptor 2 inverse agonists: design, synthesis, and binding mode prediction. *J. Med. Chem.* 52, 433–444
61 Michielan, L. *et al.* (2008) Linear and nonlinear 3D-QSAR approaches in tandem with ligand-based homology modeling as a computational strategy to depict the pyrazolo-triazolo-pyrimidine antagonists binding site of the human adenosine A2A receptor. *J. Chem. Inf. Model.* 48, 350–363
62 Li, M. *et al.* (2008) Computational studies of the binding site of alpha1A-adrenoceptor antagonists. *J. Mol. Model.* 14, 957–966
63 Moro, S. *et al.* (2006) Ligand-based homology modeling as attractive tool to inspect GPCR structural plasticity. *Curr. Pharm. Des.* 12, 2175–2185
64 Cohen, N.C. (2007) Structure-based drug design and the discovery of aliskiren (Tekturna): perseverance and creativity to overcome a R&D pipeline challenge. *Chem. Biol. Drug Des.* 70, 557–565
65 Kohring, K. *et al.* (2008) Development of benzophenone-based farnesyltransferase inhibitors as novel antimalarials. *ChemMedChem* 3, 1217–1231
66 Noronha, G. *et al.* (2007) Discovery of [7-(2,6-dichlorophenyl)-5-methylbenzo [1,2,4]triazin-3-yl]-[4-(2-pyrrolidin-1-ylethoxy)phenyl]amine – a potent, orally active Src kinase inhibitor with anti-tumor activity in preclinical assays. *Bioorg. Med. Chem. Lett.* 17, 602–608
67 Cywin, C.L. *et al.* (2007) Discovery of potent and selective PKC-theta inhibitors. *Bioorg. Med. Chem. Lett.* 17, 225–230
68 Deng, Q. *et al.* (2008) Molecular modeling aided design of nicotinic acid receptor GPR109A agonists. *Bioorg. Med. Chem. Lett.* 18, 4963–4967
69 Park, H. *et al.* (2008) Discovery of novel alpha-glucosidase inhibitors based on the virtual screening with the homology-modeled protein structure. *Bioorg. Med. Chem.* 16, 284–292
70 Park, H. *et al.* (2008) Discovery of novel Cdc25 phosphatase inhibitors with micromolar activity based on the structure-based virtual screening. *J. Med. Chem.* 51, 5533–5541
71 Hellmuth, K. *et al.* (2008) Specific inhibitors of the protein tyrosine phosphatase Shp2 identified by high-throughput docking. *Proc. Natl. Acad. Sci. U. S. A.* 105, 7275–7280
72 Cozza, G. *et al.* (2008) Identification of novel protein kinase CK1 delta (CK1delta) inhibitors through structure-based virtual screening. *Bioorg. Med. Chem. Lett.* 18, 5672–5675
73 Clem, B. *et al.* (2008) Small-molecule inhibition of 6-phosphofructo-2-kinase activity suppresses glycolytic flux and tumor growth. *Mol. Cancer Ther.* 7, 110–120
74 Diaz, P., *et al.* (submitted) 2,3-Dihydro-1-benzofuran Derivatives as a Novel Series of Potent Selective Cannabinoid Receptor 2 Agonists: Design, Synthesis, and Binding Mode Prediction through Ligand-steered Modeling.