

CURRENT PROGRESS IN STRUCTURE-BASED RATIONAL DRUG DESIGN MARKS A NEW MINDSET IN DRUG DISCOVERY

Valère Lounnas^{a,*}, Tina Ritschel^b, Jan Kelder^c, Ross McGuire^d, Robert P. Bywater^e, Nicolas Foloppe^f

Abstract: The past decade has witnessed a paradigm shift in preclinical drug discovery with structure-based drug design (SBDD) making a comeback while high-throughput screening (HTS) methods have continued to generate disappointing results. There is a deficit of information between identified hits and the many criteria that must be fulfilled in parallel to convert them into preclinical candidates that have a real chance to become a drug. This gap can be bridged by investigating the interactions between the ligands and their receptors. Accurate calculations of the free energy of binding are still elusive; however progresses were made with respect to how one may deal with the versatile role of water. A corpus of knowledge combining X-ray structures, bioinformatics and molecular modeling techniques now allows drug designers to routinely produce receptor homology models of increasing quality. These models serve as a basis to establish and validate efficient rationales used to tailor and/or screen virtual libraries with enhanced chances of obtaining hits. Many case reports of successful SBDD show how synergy can be gained from the combined use of several techniques. The role of SBDD with respect to two different classes of widely investigated pharmaceutical targets: (a) protein kinases (PK) and (b) G-protein coupled receptors (GPCR) is discussed. Throughout these examples prototypical situations covering the current possibilities and limitations of SBDD are presented.

MINI REVIEW ARTICLE

How drugs were discovered in the past

Historically, endogenous bioactive ligands were identified long before their cognate biomolecular receptor could be isolated and before their chemical structure was established [1-3]. Early information on ligand chemical structure was rudimentary and scarce compared with the many details provided by modern experimental chemistry and calculated wave functions. However, that early information often proved sufficient to infer knowledge and develop hypotheses on the physiological relevance of these ligands, even though the underlying physiological mechanisms were unknown.

The steroid hormones extracted from the bile acids at the beginning of the 20th century is an emblematic example of this paradigm. In the 1920s and 1930s several Nobel prizes were awarded for discoveries connecting the steroids with the oxidation process in living cells, vitamins, sexual development and pregnancy, as well as relating the steroid hormones produced by the adrenal cortex with biological effects [4]. This led to the recognition of the

polycyclic carbon framework of steroids. Using expert methods based on the stepwise pyrolytic breakdown of chemical structures, considerable efforts were spent to determine how the component rings of steroids were arranged. In 1932, the advent of X-ray diffraction was decisive in revealing directly the 3-dimensional arrangement of the atoms constituting the steroid scaffold, and the exact position and stereochemistry of the substituents [5]. Combining the structural details of steroidal ligands with biochemical knowledge showed how variations in the chemical groups substituted at specific positions around their scaffold were related to specific biological activity; hence allowing new medicines to be developed [4,6].

Nowadays drug discovery is driven by biological targets, genetic studies, transgenic animal models, molecular biology, gene technology, and protein science. Hence, the structure of the target biomolecule is of great practical interest. With synchrotrons and fast computers, drug designers can visualize ligands bound to their target providing a wealth of details concerning the non-bonded interactions that control the binding process (Van der Waals repulsive and attractive forces, Hydrogen-bonds, salt-bridges, and mediation by water molecules and ions). Various computational techniques (visualization, docking, molecular simulations, quantum chemistry, *de novo* design) have been devised to score the ligand-protein interactions, including attempts to estimate the binding affinity of novel molecular entities with pharmacological activity.

Despite the availability of many co-crystallized ligand-receptors X-ray structures, as well as a flora of computational methods that can be accessed via sophisticated molecular modeling software, only part of the physical reality can be perceived and/or rendered by modern computer-based techniques, casting suspicion on the overall validity of the field [7,8]. For example, the thermodynamics of the ligand-receptor association cannot be simply inferred from calculating close contact interactions, a situation which dramatically hinders scientific efforts toward truly effective rational drug design [7,8].

^aCMBI, NCMLS Radboud University, Nijmegen Medical Centre, Geert Grooteplein 26-28, 6525 GA Nijmegen, The Netherlands

^bComputational Drug Discovery, CMBI, NCMLS, Radboud University Medical Centre, Geert Grooteplein 26-28, 6525 GA Nijmegen, The Netherlands

^cBeethovengarde 97, 5344 CD Oss, The Netherlands

^dBioAxis Research BV, Pivot Park, Molenstraat 110, 5342 CC Oss, The Netherlands

^eMagdalen College, High Street, Oxford OX1 4AU, United Kingdom

^f51 Natal Road, Cambridge, United KingdomUK

* Corresponding author. Tel.: +33 490221789

E-mail address: v-lounnas@unicancer.fr (Valère Lounnas)

Phase diagram for pre-clinical drug discovery

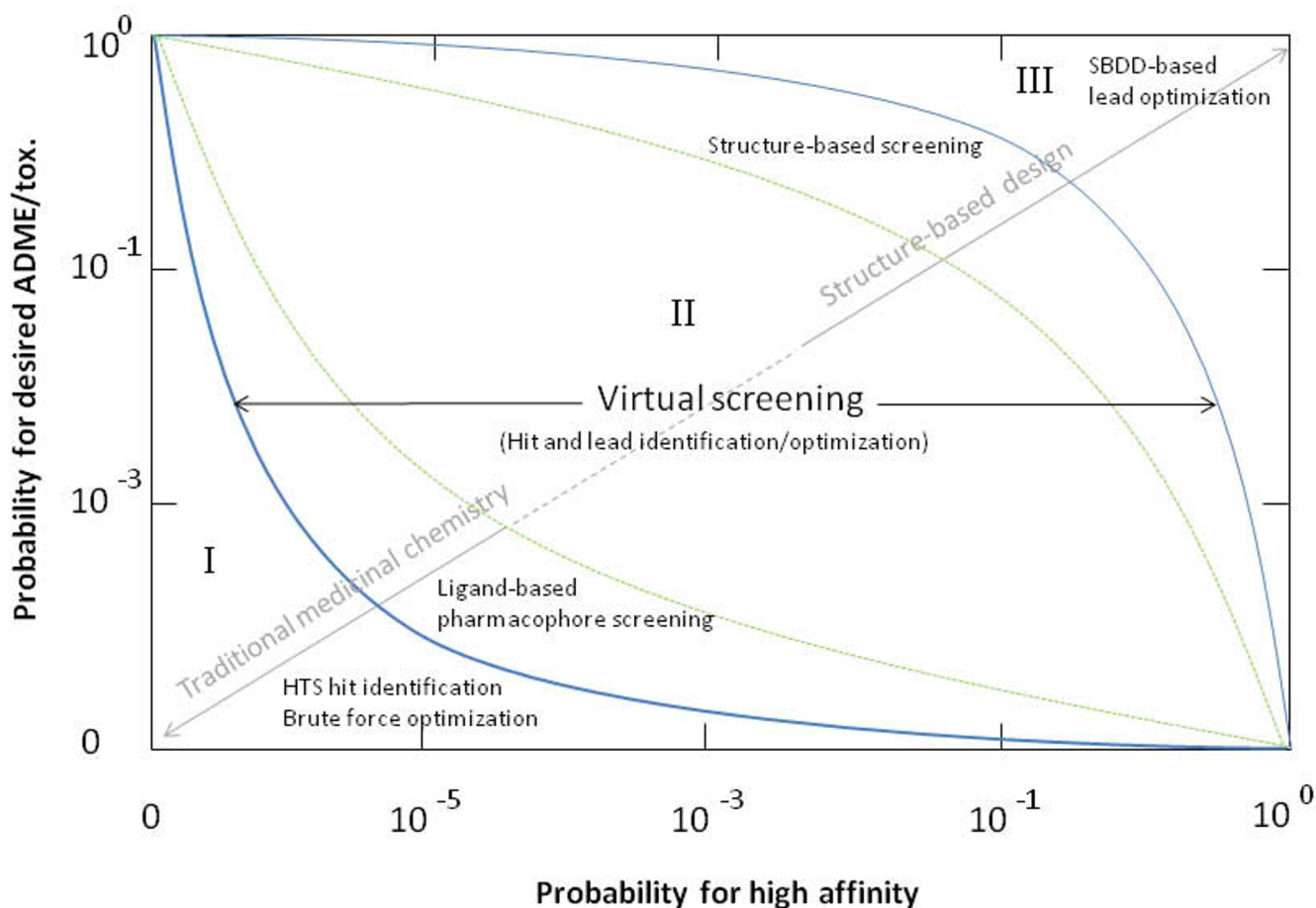


Figure 1. Idealized description of the probability landscape of the pre-clinical drug discovery process. From the identification of hit compounds with low to medium affinities to the optimization of lead compounds with high affinity and favorable ADME/tox, the process of preclinical drug development is a highly complex problem. Affinity and desired ADME/tox constitute two independent (orthogonal) dimensions that, unfortunately, cannot be easily optimized separately. Classical/brute force high-throughput identification and optimization of hits too often fails to provide compounds with both high affinity and adequate ADME/tox properties. This situation confines project outcome to a low-success probability pitfall represented by the bottom of the phase diagram (area I). Escape from this zone can be achieved with the help of a broad variety of virtual screening techniques encompassing pharmacophore-based techniques and structure-based techniques (area II). Ultimately, structure-based considerations can help drug designers guide a project towards a more productive area (III) of the phase diagram. Orders of magnitude for probabilities are only a rough guess and the frontiers between areas are schematic.

In this minireview we examine the methodological trends that have emerged recently in the computer-aided molecular design of pharmacologically relevant ligands and how successful attempts were made to rationally combine X-ray, modeling and calculation techniques.

The landscape of modern drug discovery

The paradigm of probabilities in drug discovery

Are there more stars in the universe than possible organic molecules with a molecular weight < 600? Observation of the cosmos has led astrophysicists to map the universe and suggest that there are about 10^{23} stars gathered in 10^{11} galaxies [9]. In parallel, consideration of the real number of possible ligands has been the subject of savvy estimates [10,11]. Complicating the matter is the fact that not all chemically plausible molecular structures might be synthetically accessible nor might they be affordable. Nevertheless, figures commensurate with the number of stars have been proposed. Neither the universe nor the ensemble of possible ligands can be explored

systematically. The surrounding universe hints that extraterrestrial civilizations may exist, but the odds that a spaceship travelling for centuries at the speed of light may encounter one of them would remain so small that such enterprise would be doomed; a dreadful calculation which most science fiction aficionados are unaware of. A similar challenge is faced by high through put screening (HTS) widely used by the pharmaceutical industry in hit compound identification. A spokesperson from the industry modestly acknowledges this issue, skillfully admitting that: 'the discovery task has shifted somewhat during the past few years from simply identifying promising leads to the added proviso that dead-end leads should be eliminated from consideration as early in the process as possible' [12,13]. In other words, the leads supplied by classical chemistry optimization rounds performed around HTS-supplied hits are often difficult to transform into drugs. Optimization of binding affinity in isolation by traditional medicinal chemistry methods leads to poor ADME/tox properties through effects such as the inclusion of bulk to ligands, excessive functionalization, growth of hydrophobic groups and/or selection of functional groups with supposedly known ADME/tox liabilities [14]. They are too few leads, they are not diverse and more importantly,

HTS provides no information at all about the way they interact with the target receptor, thereby precluding efficient optimization (Figure 1). The screened collection tends to represent what happens to be available in a particular organization, instead of rationally selected chemotypes. HTS favours quantity over quality and insight, and results in large amounts of data of dubious quality, which requires much time and effort to be analyzed. One could argue that the vast resources invested in HTS could have been used more productively to characterize how leads interact with the target receptor, to inform their efficient optimization.

In response, many research groups throughout the world have dedicated efforts toward structure based drug design (SBDD) and virtual ligand screening (VLS), as the most scientifically promising approaches to identify ligands for pharmaceutical targets [15-20]. SBDD provides insight in the interaction of a specific protein-ligand pair, allowing medicinal chemists to devise highly accurate chemical modifications around the ligand scaffold, whereas the strength and versatility of VLS are self-evident as it can be used to filter compounds to be assayed based on any rationale that can be expressed in terms of molecular properties (pharmacophores and/or physico-chemical descriptors). Here, we should point out that docking programs can be used to filter libraries of compounds [21]. However, it remains extremely complicated to rationally analyze docking results for libraries containing chemically diverse compounds [21]. The lack of reliable of scoring functions across heterogeneous chemical classes and the frequent impossibility to predict the correct binding modes has generally cast doubt over docking as a genuine screening method. We shall see in further sections how one may combine VLS with SBDD and how docking can be efficiently used as a virtual screening tool as well in specific circumstances. SBDD offers distinctive advantages compared with the classical ligand-based techniques such as the comparative molecular field analysis (CoMFA) and pharmacophore-based techniques, based on only ligand information, that have well known limitations [22]. Another crucial advantage is that electronic libraries can cover already existing compounds as well as compounds that have not been synthesized yet but are chemically relevant.

SBDD can be broadly envisioned as one intrinsic component of rational approaches based on screening assays that are more adapted to drug discovery than HTS. For instance, fragment-based drug design (FBDD) has emerged as a very powerful means to reverse the HTS probability paradigm by exploiting molecular fragments with molecular weight < 150-250 [23,24]. These fragments can be selected based on adequate physico-chemical properties, collected in libraries, assayed and, using SBDD-derived information, reassembled into larger molecules with improved drug like characteristics [25-27]. FBDD is a genuine branch of rational drug design for which all the principles we discuss in this minireview are valid and can be used. We refer the reader to recent reviews on FBDD [28,29,30].

SBDD can reveal important information on the structure activity relationships of a chemical series, especially in the lead optimization phase, when very accurate modifications are needed to adjust an ADME/tox profile while maintaining binding affinity. Target selectivity is another issue where SBDD may be vital, especially when dealing with large classes of pharmaceutical targets SBDD requires, however, that an X-ray structure be available for the pharmaceutical target, or that a 'reasonably' valid homology model has been constructed. We shall illustrate in the next sections what reasonable means. Knowledge of the target structure plays the same role as boundary conditions in physical sciences, providing indications for instance on the maximum possible volume that the ligand can occupy, and the particular loci and orientations where hydrophobic and/or hydrophilic interactions can be engaged.

Is there a missing link in SBDD?

Ligand-based approaches (e.g., CoMFA and pharmacophore-based methods) are intended to capture and/or exploit a remarkable pattern detected predominantly from the empirical superposition of a set of experimentally identified binder and non-binder compounds [31,32]. In this respect, only the first-order/gross characteristics of the ligand-receptor recognition and binding pattern are taken in consideration. With these techniques virtual library screening tends to retrieve large sets of compounds (hundreds to thousands) that are usually further filtered based on empirical considerations, excluding compounds that are not drug-like and relying on 'luck' (see previous section) to identify new binders within the remaining compounds.

In contrast SBDD aims to provide a richer picture of the ligand receptor interactions. However, SBDD has developed more slowly. Many X-ray structures of pharmaceutical targets co-crystallized with their natural ligands and/or inhibitors are publically available, however many more have been resolved by the pharmaceutical industry without being published [22]. Thus, it is difficult to estimate the extent to which these data have contributed in conceiving new drugs. This situation has hindered a fair evaluation of the true potential of SBDD by the scientific community and the merit of SBDD may not seem as widely praised as the ligand-based techniques. In many instances, at conferences and during informal discussions among experts, the benefit of SBDD in drug design remains a controversial issue [7]. For many drug designers, the lack of reliability in predicting affinities constitutes what could be referred to as the 'missing link' which undermines effective SBDD. But, if we do really believe that is the problem then we should simply resign as rational drug designers as there is no obvious way the accuracy gap will be bridged [33].

Intuitively, the difficulty of establishing the limit of applicability of the methods maybe more troubling to the productive development of SBDD than inaccuracy in energetic calculations. The 25th anniversary issue of JCAMD has published a series of articles that all describe one or several aspects related to this concept. Quoting one of them seems appropriate: 'Like poor mule drivers, we have a tendency to push our methods beyond their limits' [7]. To address this misbehavior we have to realize that the time scale required by rational SBDD contrasts with the state of permanent urgency that characterizes the pharmaceutical industry in its quest for introducing new drugs to the market [34]. In this setting, a lead optimization program hardly lasts 12 to 18 months, a time too short to gather all experimental information and model validation necessary for detailed SBDD. Pharmacophore-based techniques are well adapted to this limit since they require far smaller time overhead than SBDD to establish a validated albeit crude model of receptor-ligand interaction [35]. But the rewards of SBDD are potentially greater compared with those techniques that have proven their paucity in providing sufficiently accurate rationales, especially in the lead optimization phase where not one but several properties must be optimized together.

We need to keep an eye on the components of the free-energies of binding!

Despite considerable progress achieved in structure-based virtual ligand screening and computer-aided drug design [36-43], the reliable and practical prediction of binding affinities remains an elusive holy grail [44-51]. Even when knowing the binding mode of a ligand to its receptor, its binding affinity cannot be inferred reliably although, paradoxically, it can be measured highly accurately with titration calorimetry and surface plasmon resonance [52-54].

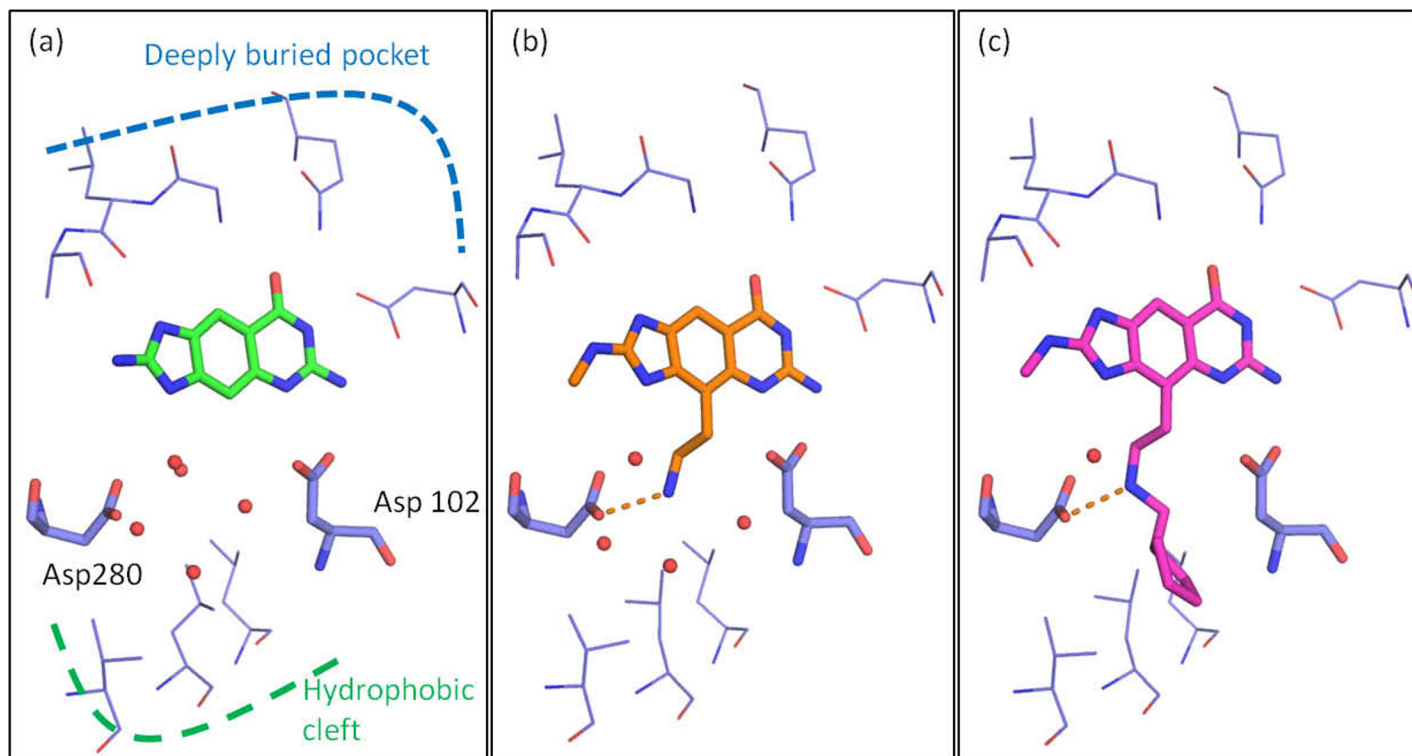


Figure 2. Example of how SBDD can be applied to displace highly ordered (immobilized) waters in the active site of an enzyme: (a) X-ray structure of *lin*-benzogunaine in complex with TGT with a binding affinity 58 nM; (b) an amine substituent that could mimic the water molecules while maintaining the same level of affinity for TGT (55 nM) was introduced; (c) substitution of the amine to occupy the small hydrophobic pocket with a cyclohexyl moiety increased the binding affinity significantly (2nM).

The affinity of the ligand for its receptor is dictated not only by the enthalpic contribution arising from close range physical interactions (Van der Waals interaction, hydrogen bonds, ionic pairing) but also by an entropic contribution that is a measure of the overall system disorder. In biology though, the concept of disordered systems is difficult to comprehend and is sometimes counter intuitive, as a ligand bound to its receptor may seem inherently a more ordered state than the unbound ligand and receptor when this is not necessarily the case [8]. Indeed, the entropy penalty resulting solely from the reduction of the overall body translational and rotational freedom of the ligand and receptor may amount up to 15 kcal/mol while the entropy penalty due to the reduction of rotational freedom is estimated to be ca. 0.6 kcal/mol per rotatable bond [53,55]. Upon binding, favorable interactions with the receptor compete with these unfavorable energy terms. In addition, the disruption of ordered clusters of water molecules around the ligand hydrophobic moieties, a process which causes a gain of disorder in the solvent, contributes favorably to the stability of the ligand-receptor association [56,57]. Experimental measurement showed that increasing the nonpolar surface area of an aliphatic alcohol by inserting a CH₂ group has an unfavorable entropic contribution of +1.7 kcal.mol⁻¹ to the free energy of transfer in aqueous solution, which is only partly compensated by an enthalpic contribution of -1.3 kcal.mol⁻¹ [58]. The result is in an overall unfavorable free energy of 0.4 kcal.mol⁻¹. If not counterbalanced by other effects, a change of this magnitude in the hydration energetics of a ligand corresponds to an increase by a factor of 2 in its measured binding affinity (IC₅₀); and we know that optimization of a lead often operates in a range less than 1.4 kcal mol⁻¹ (i.e., less than a factor of 10 in IC₅₀s) [7]. Increasing the hydrophobicity of a ligand is a strategy often used in drug discovery. However, this change has to be accommodated by the receptor and

the ligand must remain a potential drug; ligands with excessively increased hydrophobicity have poor absorption, distribution and off-target toxic binding properties [14,59]. Efforts to gain insight on how different chemical groups branched around a same scaffold may influence binding affinity are hampered because the experimental evaluation of the enthalpy (ΔH) and entropy ($T\Delta S$) of binding (with ITC) are large quantities that compensate each other while having error bars exceeding 1.4 kcal mol⁻¹, which is uncomfortably large in comparison to the overall binding energies [60]. In standard assays, however, the error in the binding free energy is usually in the range +/- [0.1 - 0.5] kcal/mol only, allowing results to be more readily exploited in drug design, though deeper insight in the binding process is limited.

An often neglected aspect in rational drug design is the ubiquitous presence of water inside and around the receptor binding cavity and in some instances, the free-energy cost of displacing ordered or partially ordered water molecules involved in the residual solvation of active site residues should be considered carefully [61-64]. Computational methods are being devised to calculate the energy cost of displacing clusters of rotationally and translationally hindered water molecules present in binding site cavities accessible to water [65-70].

When a ligand binds to a protein, water molecules surrounding the ligand and protein need to rearrange. These rearrangements contribute to the binding affinity of the ligand. Most water molecules building the hydration shell around the ligand and in the protein active site need to be either removed or displaced. Paying attention to the thermodynamics of these phenomena, for which no relevant experimental data are available, has been done in early theoretical studies of protein folding and protein association [71]. Very recently, molecular dynamics simulation (MDS) has further contributed to our understanding by showing that water molecules in the first hydration

layer around a receptor surface are less thermodynamically stable than bulk water [61]. The process of transferring water from the bound state to bulk solvent is associated with a favorable free energy [61,64,65]. This, however, does not hold for water molecules that are deeply buried and tightly bound in crevasses at the surface or inside the active site of proteins. The maximum entropy change of releasing such a water molecule into bulk water is evaluated around 2 kcal.mol⁻¹; a gain which is unfavorably outweighed by the loss of enthalpy due to the breaking of hydrogen bonds with the receptor [63]. Because they are highly ordered (immobilized), these water molecules are seen in X-ray crystal structures, and in many cases ligand binding is often mediated by such water molecules [66]. Attempting to displace them with experimental ligands may not only be difficult to achieve but also have unpredictable unfavorable effects [63,72,73].

However, when a ligand is binding to a hydrophobic cleft at the surface or inside the protein the enthalpic contribution to the free energy of binding will increase because of favorable interactions between the ligand and protein. Simultaneously, water molecules that occupy the hydrophobic cleft will be released into the solvent, resulting in a favorable increase in entropy [61].

The discovery and optimization of inhibitors for the enzyme tRNA-guanine-transglycosylase (TGT), an enzyme from pathogenic *Shigella* bacteria, is an example where desolvation of the ligand and protein has a significant influence on the binding affinity of the inhibitors [62,74]. The tRNA binding site of TGT consists of one deeply buried pocket surrounded by smaller, solvent exposed pockets (Figure 2) in between two negatively charged catalytic aspartates. Using classical SBDD approaches, attempts were made to substitute the *lin*-benzoguanine scaffold with a scaffold that occupies the hydrophobic cleft situated below the two aspartates to improve potency. However, occupation of this hydrophobic cleft did not result in the expected gain in binding affinity.

In the crystal structure of *lin*-benzoguanine in complex with TGT, an ordered solvation shell solvating the two proximal charged aspartates was discovered [75]. Functional groups such as amine substituents that could mimic the water molecules were introduced. Comparing the binding affinity of the scaffold *lin*-benzoguanine (Figure 2a, 58 nM) with the unsubstituted amine (Figure 2b, 55 nM), binding affinity could be maintained (cost neutral) while restoring the favorable interaction provided by the displaced network of water hydrating the aspartates. Hydrophobic substitutions on the amine (with for example a cyclohexyl moiety) increased the binding affinity significantly by accessing a lipophilic pocket below the catalytic aspartates (Figure 2c, 2nM).

Homology models to generate and verify binding mode hypothesis

The choice of the template(s) used for the homology model construction may lead to C α positional variations of the order of 1 to 3 Å (or more) in the active sites of enzymes, depending on the percentage of sequence identity shared with the elected template(s) [76,77]. Such variations have to be appraised during a model construction because they may lead to substantial side chain displacements in the final model. Various techniques can be used to mitigate this problem: the use of multiple templates or a mixed template etc. In principle, homology modeling software can be used to optimize side chain conformations and the non-bonded interactions between them by simulated annealing, conditional probability distribution, or potential of mean force [78]. At this stage there are no safe recipes to apply but careful physical insight along with verification that hydrogen bond networks and hydrophobic residue packing statistically conform with the knowledge-based characteristics

extracted from the X-ray structures [79]. Conserved interactions between particular residues (hydrogen bonds, ionic bridges,...) in the active site may be also used as a guide to adjust the model during its elaboration [80,81]. One has to avoid the pitfall of model overoptimization that might actually drive the model away from physical reality. At the same time the model should be consistent with respect to most (if not all) available experimental data [77,79]. Failing to abide by this necessity inevitably leads to a high risk of project failure in a later phase when the simultaneous optimization of other properties, such as absorption, metabolism, excretion and toxicity (ADME/tox), must be performed for a lead compound to become a real drug, while retaining a high affinity and selectivity for the pharmacological target. The next sections illustrate with recent examples that this approach can be effectively put in practice; successfully combining two or more modeling techniques together (Homology Modeling, Docking, MDS,...).

Different types of ligand-receptor complexes and different challenges

There are a plethora of X-ray structures of receptor-ligand complexes for several important families of drug targets, including proteases, HIV reverse transcriptase, nuclear receptors (NR), protein kinases (PK), and even RNA targets [82]. Structures have also started to become available for some G-protein coupled receptors (GPCR). Some of these receptors are already privileged targets for SBDD (e.g., NR, PK) and some others might also provide other avenues for SBDD in due course (e.g., GPCR, RNA) [83,84]. This wealth of structural information will continue to increase, also as a result of the structural genomics initiatives [85].

When exploiting structural information, drug designers are confronted with situations that bear similarities within the same class of targets. This gives rise to what we refer to as prototypical situations of ligand-receptor recognition, but not simply 'routine situations' because each project requires its own logical development. Different types of receptors possess unique challenges for SBDD. We present two example systems here:

- ⇒ Protein kinases (PK) (Figure 3): multidomain proteins with complex catalytic domain ligand binding sites (large cleft between two domains); ATP is the natural ligand; weakly bound water and Mg or Mn ions mediate ATP binding; the vast majority of PK inhibitors target the ATP binding site; selectivity is a crucial issue that requires SBDD-based considerations; a well defined purine binding motif acts as an anchor; possibility of allosteric binding/inhibition; activation occurs via loop phosphorylation; many crystal structures of complexes exist and homology models can be routinely constructed; PKs are well suited for SBDD.
- ⇒ GPCR (Figure 4): 7-helix bundle transmembrane assembly; very versatile binding modes, deeply buried inside the transmembrane region; possibility of allosteric binding, agonism and antagonism effects, selectivity is an issue; binding cavity is very complex and topped by large highly flexible loops; mechanism of activation/inhibition requires more in depth understanding; not yet routinely amenable to SBDD; more X-ray structures and binding experiments combined with mutational data are expected to clarify this mechanism.

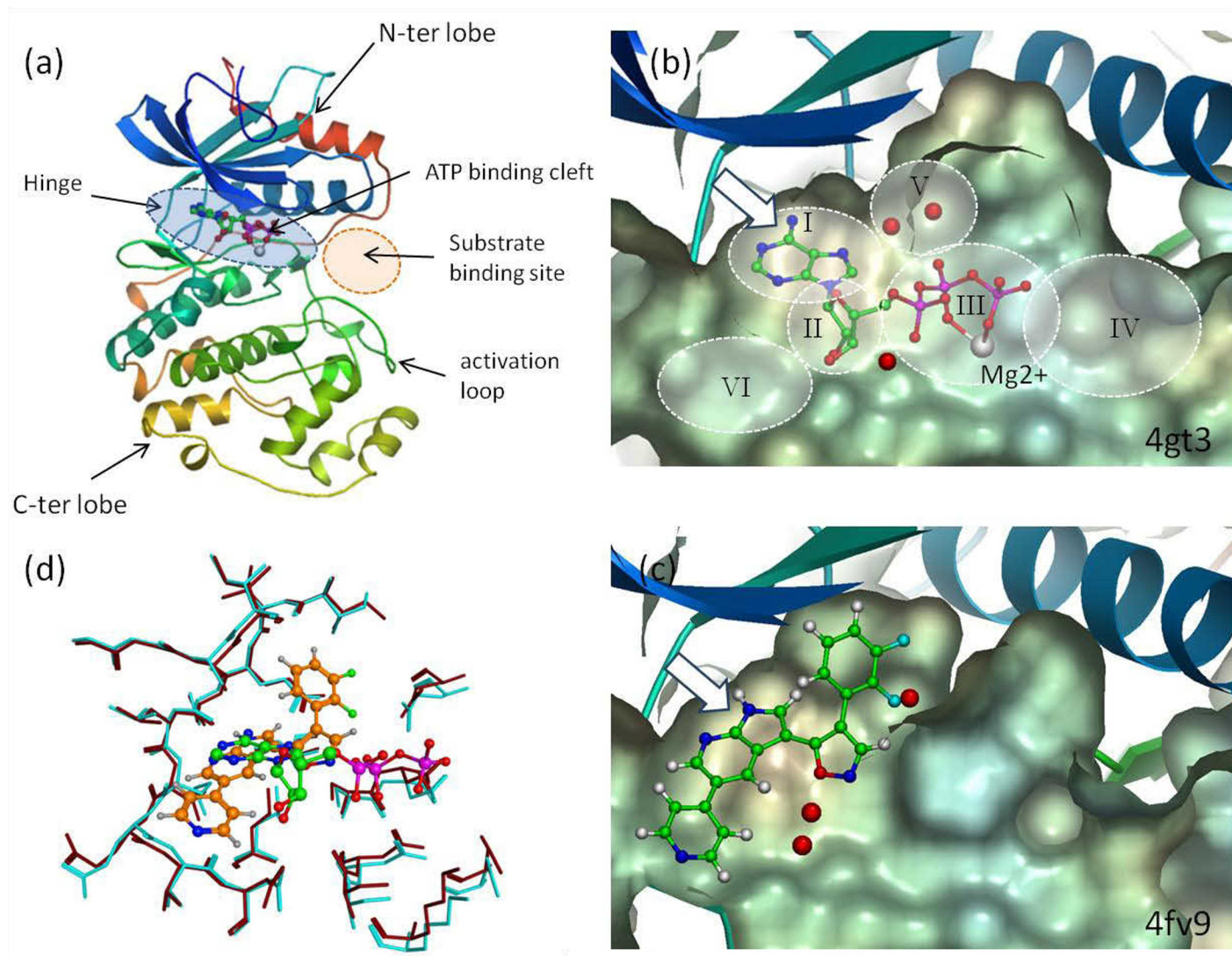


Figure 3. (a) Ribbon representation of Erk2 kinase in its activated form with ATP bound (pdb entry 4gt3). The ATP binding cleft has a large solvent exposure and can be decomposed into several compartments or pockets (I to VI)(b). The average rate of conservation for the residues lining it is only 51% [80]. Together these observations allow the design of a large variety of selective inhibitors. Subareas (b) and (d) show clipped views of the ATP binding site seen from the top, for two X-ray structures of Erk2 with, respectively, ATP (pdb entry 4gt3) and the synthetic ligand E71 (pdb entry 4fv9). Note that highly ordered water molecules (red spheres) mediate the binding in both cases. The synthetic ligand is anchored to the hinge residues via the same two hydrogen bonds (bidentate motif; white arrow) as does ATP. However, it is oriented differently and expands perpendicularly to the cleft main axis, occupying the buried pocket (V) on one side and extending to the outer part of the cleft (VI) on the other side (c). Despite considerable difference in the chemical structures and binding modes of ATP and E71 the deviation between the backbone and side chains atoms of the two X-ray structures is small ($<1\text{\AA}$) (4fv9 brown and 4gt3 cyan/blue) (d). Although it is not true in 100% of cases, conformational invariance of the active site residues can be reasonably assumed in many SBDD-based drug design projects.

Designing selective ATP competitors to modulate protein kinase activity

Although PKs constitute only 2% of the human genomes they represent $\sim 25\%$ of putative drug targets [86,87]. They are multidomain proteins that intervene in many biological processes, especially signal transduction. Their catalytic domain (called kinase domain) chemically modifies other proteins by transferring the γ -phosphate group from a nucleoside triphosphate (usually ATP) to the hydroxyl group of specific residues such as serine, tyrosine, threonine, and sometimes to histidine. Deregulation of these processes is implicated in many physiological disorders, including inflammatory diseases and cancer [88,89]. Hence, protein kinases are prominent drug targets as reflected by the number of deposited patents [90].

Since the first X-ray structures of the kinase domain appeared, it was understood that the kinase domain had a highly conserved fold across all members of the superfamily [91-93]. This fold comprises two connected domains that form a cleft where ATP binds, adjacent to a shallow substrate binding region (Figure 3). The fact that proteins can be categorized in a large superfamily sharing a common fold has far-reaching practical impact in drug design because it is expected that a substantial amount of information gleaned from one kinase can be transferred to other ones. Indeed, the conserved natural ligand (ATP) of PKs means that their binding site is also somewhat structurally conserved, which was initially perceived as a drawback preventing the discovery of selective ATP-competitive inhibitors. Competing with ATP was also deemed difficult since it is abundant in cells [94].

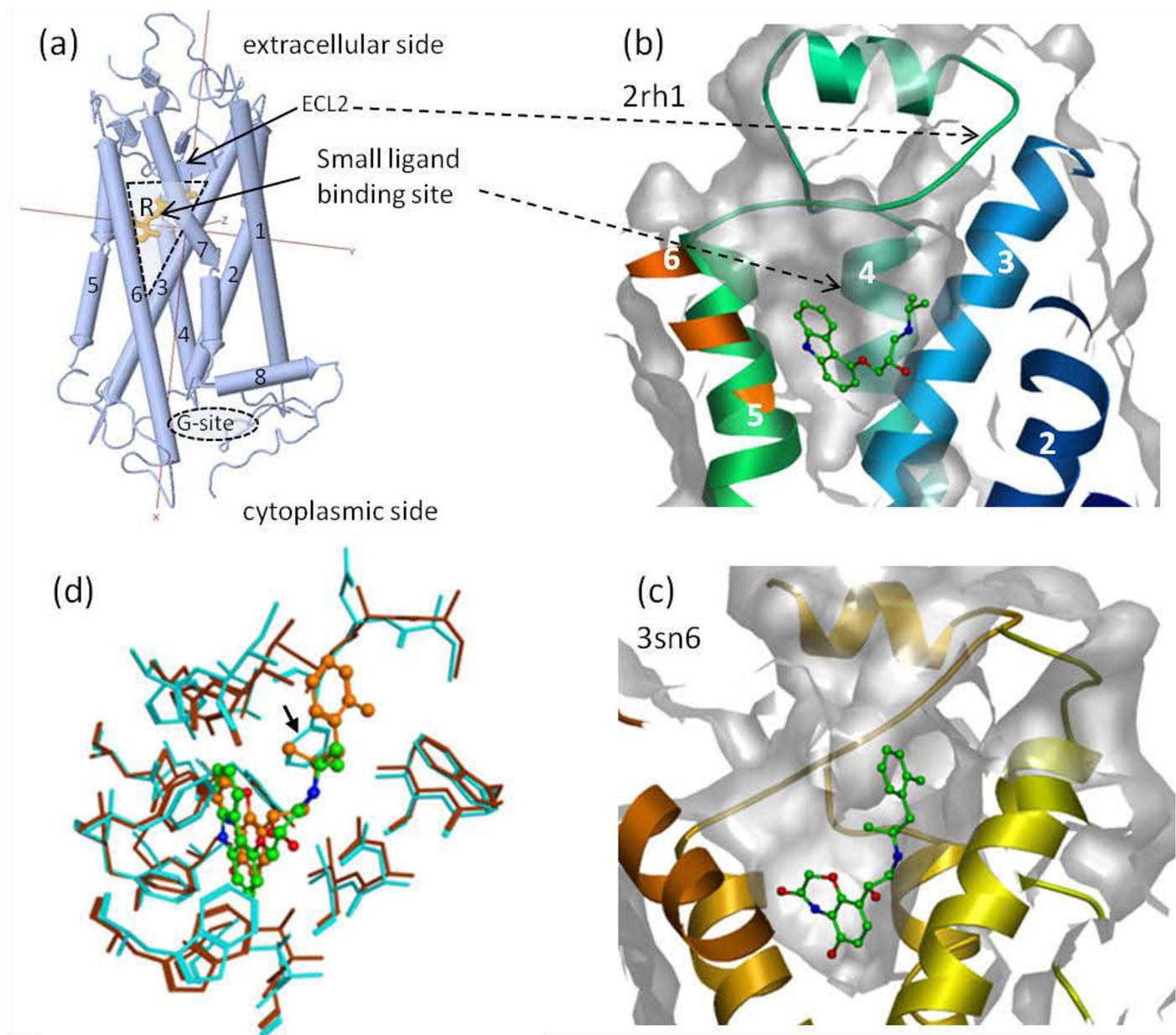


Figure 4. (a) Cartoon representation of X-ray structures of bovine rhodopsin, representative of class A GPCRs: with retinal (R) in the cis form bound before it has absorbed light (inactivated form, PDB* entry 1u19; side view (a)). GPCRs are made of a bundle of 7 alpha helices (numbered from 1 to 7 in view (a) and (b)) topped by a large β -pleated sheet structure resulting from the packing of two helix connecting loops. One additional helix (8) located at the interface between the cell membrane and cytosol has a regulatory role in deactivating the receptor. A major evolutionary constraint on GPCR function is to bind G-proteins in a large crevasse at their cytoplasmic side (G-site) [128]. The structural change that leads to GPCRs activation (G-protein coupling) is a general mechanism controlled by an extremely broad array of endogenous effectors (modulators) ranging from large structured peptides such as chemokines and peptide hormones to low molecular weight ligands that bind in the cone-shape cavity buried inside the 7-helix transmembrane bundle near the extracellular side of the receptor. Illustration of GPCR plasticity is given with: (b) zoomed side view of X-ray structure of β 2AR in complex with the antagonist (S)-Carazolol (β 2AR inactive state, PDB entry 2rh1); (c) X-ray structure of β 2AR in complex with a high affinity agonist (BI-167107) (β 2AR-Gs protein active complex, PDB entry 3sn6); (d) overlay of the side views (b) and (c) showing residues in close contact (< 5 Å) with the antagonist (cyanide blue) and the agonist (brown). Positional shifts of side-chain and main-chain atoms highlight the plasticity of the small ligand-binding site; the side chain of Phe 193 (black arrow) was too disordered to be resolved in the agonist bound structure. (*see reference [82])

On the other hand, this situation has the advantage that structural models may be obtained by homology (comparative) modeling for most kinase domains [80]. This should signify a successful future for SBDD, as long as, modelers assess the potential weakness of their model prior to its use and bear in mind a number of practical limitations that continue to hamper the development of protein models constructed by homology modeling. First, the opening of the two lobes that sandwich the ATP binding site can vary by a few

degrees, due to the flexibility of the hinge region connecting them, causing substantial displacements (up to several Å) in the relative positioning of the residues lining ATP at the other extremity of the binding pocket making the ideal alignment of structures often ambiguous [92,95]. Second, as happens almost systematically in homology model construction, the multiple sequence alignments used for selecting templates cannot be 100% ascertained because of gaps and insertions, especially with superfamily members characterized by a

low sequence identity (< 40%), even around the active site, and even when multiple templates are used and class-specific information (position-specific sequence conservation) is implemented in the sequence alignment [76,96,97]. Third, despite elaborate algorithms using conformational propensities derived from X-ray structures, it is difficult to accurately assert which residue side chain conformations are physically relevant for each particular ligand-receptor association and the positions of the residue side chain lining the active site of a homology constructed model remains questionable [79,98-100]. Fourth, the possible effect due to the presence of mediating water molecules and/or counter ions inside the active site is difficult to predict, except for the few conserved (tightly bound/structural) examples that are inferred from X-ray structures [101]. Many research groups have developed *state-of-the-art* semi-automated methods that allow experienced drug designers to construct knowledge-based homology models where issues related to limitations 1, 2 and 3 can be tackled in a 'fairly' rational way in many cases and, less frequently, in an evidence-based way [79]. However, the way these models are built cannot be generalized and their validity must be assessed on a per individual project basis due to the conditions and specific interactions involved in ligand binding being extremely variable.

A kinase structural database (Kinase Knowledge Base, KKB) has been recently developed that covers about 148 different kinase domain structures that have been deposited in the Protein Data Bank (PDB) and which represents about 20% of all kinase domains in the human genome [80]. Using a carefully selected set of 20 non-redundant X-ray structures overlaid with a protein structure alignment tool a structure-based sequence alignment of the 594 members of the human kinome has been generated [102].

This work opens the possibility for drug designers to benchmark their homology model construction methods. For instance, they can select a dozen kinases out of the 148 X-ray structures and reconstruct them by homology based on the kinome-wide sequence alignment provided in the KKB database. The comparison between the reconstructed set and the X-ray structures can provide drug designers valuable feedback to test their methodology.

The analysis of sequence conservation for active site residues shows the residue variability and character (hydrophobic, flexible, polar, aromatic, positively or negatively charged, and mixed) at every particular position in the ATP binding pocket [80]. This high variability together with the finding of kinase inhibitors that have a high affinity for a relatively reduced subset of kinases enables the SBDD approach (Figure 3) [102]. We shall show how in the detailed analysis presented in the next sections.

The combined techniques approach

Docking-based spatial superposition of ligands to produce synergy with QSAR

Accurately conducted docking studies can be used very efficiently to complement QSAR techniques such as CoMFA and comparative molecular similarity index analysis (CoMSIA) which crucially rely on how compounds are overlaid [103]. However, the validity of the docking methodology must be established by successfully reproducing the binding modes of some of the ligands belonging to the set of studied derivatives and for which X-ray structures are available in the PDB. The remaining compounds are docked following the same docking protocol. The purpose of docking is to provide a rationally derived structural alignment (spatial superposition) of the studied set of compounds, that in turn can be used to derive a predictive QSAR model based on molecular descriptors including steric, electrostatic,

hydrophobic, and hydrogen bond donor fields as well as PSA and MLogP.

We see how a multi-stage approach and the use of two computational methods can complement each other and compensate for the respective weaknesses of each method, thus creating a synergy in SBDD. Docking scores are unreliable, and most QSAR techniques cannot be applied without either knowledge or a strong hypothesis of the bound conformations of the ligands and spatial alignment [104-106]. But if the preferred binding mode of chemically related compounds can be inferred, docking can provide the correct ligand conformation and an accurate structural alignment that is required to build a predictive QSAR model [107-109].

Scaffold-based virtual ligand screening

Acquiring accurate knowledge of the binding modes for identified chemical classes of compounds is such a stringent issue in SBDD that strategies have developed to reverse engineer the problem and try to identify proto-ligand scaffolds with adequate affinity and physicochemical properties selected from the molecular characteristics of known ligands. This approach requires the assay of specific subsets of compounds extracted from specialized libraries containing relatively large molecular fragments within a molecular weight range of 150 – 250 Da. For instance in the case of protein kinases, prototypical scaffolds targeting the ATP binding site are preferably flat and contain at least two conjugated ring systems which can engage bidentate hydrogen bonding interactions [18,110,111].

Once proto-ligand scaffolds that compete with ATP have been identified in biochemical assays, they can be co-crystallized with the target enzyme and their binding mode determined by X-ray crystallography. They are particularly well suited for SBDD. Virtual libraries can be subsequently constructed by identifying the substitution positions best suited to expand the ligand around the scaffolds. Docking can be used to select the putative ligands that best fit the identified binding modes and that provide additional complementarity with the residues lining the binding site [111].

This approach is particularly well suited for PKs because the ATP binding cleft is large and open which allows for multiple possibilities to expand the ligand scaffold in various directions to search for selective interactions (Figure 3) [111]. The concept of scaffold-based series can be implemented in SBDD projects with a great deal of variation and prior screening in biological assays might not always appear necessary. For instance, a docking study performed on a series of PKs has shown that docking large sets of ligands that share the same scaffold is statistically more likely to retrieve the correct class-specific binding mode within the top clusters of docked poses than clustering 100 poses generated from a single ligand [112].

SBDD with GPCRs

GPCRs (Figure 4) are important pharmaceutical targets for the treatment of many physiological disorders including some inflammatory diseases, neurotransmission and sensory perception impairment, diabetes, cardiovascular defects, AIDS and some cancers. At least 40% of the currently marketed drugs target GPCRs [113]. In the past decade, groundbreaking X-ray structures have been solved for an increasing number of class A GPCRs. The first one was the bovine rhodopsin receptor X-ray structure that has served from that time as a useful template for homology modeling of other GPCRs to replace the old bacteriorhodopsin based models. Nowadays X-ray structures are available for the adrenergic β 1A and β 2A, histamine H1, adenosine A2A, dopamine D3, muscarinic M2 and M3, opioid kappa and mu, CXCR4 receptors, etc., most bound to low molecular weight

antagonists. These structures have helped assess the many 'common mistakes' that have crippled early GPCR models [114,115]. It is fantastic that the importance of all this work starting from molecular cloning to X-ray structure was recognized by a Nobel Prize awarded in 2012 for the X-ray structure of the β 2-AR-Gs protein complex [116,117]. But the early models had served already as valuable 3D alignments that stimulated our thoughts on how ligands might bind, and have led to many successful suggestions for site-directed mutagenesis experiments to study ligand binding. Implementing valid SBDD considerations within drug design projects targeting GPCRs still remains difficult due to limited insights as to how ligands interact with them and modulate their activity [118].

Many experimental observations for rhodopsin have shown that residues from the second extracellular loop (ECL2) and located in the cone-shape cavity inside GPCRs interact with ligands to trigger the outward tilt of transmembrane helix 6 (TM6) relative to TM3 and TM5 [119]. This movement induces a concerted rearrangement of the whole helix bundle that conveys the activation signal to the intracellular side of the receptor, allowing G-protein coupling to the activated form of the receptor (Figure 4) [120-122]. This mechanism is thought to be general across class A GPCRs [123]. Both the activated and inactivated states are assumed to exist *in vivo* as an equilibrium governed by specific sets of amino acids acting as 'micro-switches' and conserved within class A subfamilies [124-128]. By interacting with these residues, agonists and antagonists stabilize one state or the other. Although the mechanism is general, it is modulated by different amino acids across subfamilies and subfamily members [129]. In 2007, the X-ray structure of the β 2-adrenergic receptor proved that the spatial arrangement of the transmembrane helices relative to each other is highly similar to that of bovine rhodopsin; which was encouraging for homology modelling and SBDD [130]. However, the conformation of ECL2 was entirely different between the two structures.

Compared with other classes of receptors such as PKs and NRs the number of elucidated X-ray structures of GPCRs bound to small compounds is still too small to safely extrapolate to other classes of ligands in SBDD [131]. However, the number of GPCR X-ray structures is growing rapidly. Already, they have helped validate the use of docking and virtual ligand-screening (VLS) applied to GPCR crystal structures [131-133]. Until recently, the GPCR conformational plasticity was thought an obstacle for SBDD of agonists [134,135]. Several X-ray structures of agonist-bound complexes have been produced in 2011 for β 1AR, β 2AR and A_{2A} AR and more structures are expected in the near future. These structures will provide an improved 3D framework for the rational design of agonists and antagonists, illustrating the utility of modeling broader structural changes than only those limited to the residues lining the binding site pocket [120,134].

The elucidation of the X-ray structure of bovine rhodopsin in 2000 put GPCR homology modelling on a reasonable footing, assuming a high overall structural conservation across members of the superfamily despite their very low sequence identities [136]. GPCR homology models of improved quality can now be prepared routinely thanks to database such as GPCRDB and GPCR-SSFE constructed in conjunction with the structural elucidation of several other class A GPCRs [114,134,135-140].

So far, antagonist/agonist ligands have been mostly identified using ligand based approaches consisting of filtering libraries of commercially available compounds with pharmacophore models [141-143]. Nevertheless, carefully constructed homology models of rhodopsin-like GPCRs are increasingly reported, with the concomitant use of docking, which allowed enrichment of screened

compounds with actives with rates similar to those obtained with the ligand-based pharmacophore approaches [144-154]. Crucially, the advantage of structure-based screening would not be only in an improved success rate but also in providing: 'useful information on the structurally vital residues involved in binding, and their unexplored potential for the future development of potent receptor antagonists' [145].

Hence, the design of GPCR modulators will ultimately require the ability to make a ligand have a defined functional profile in terms of agonism/antagonism. This assumes that the micro-switch residues with which ligands engage are correctly identified and that these interactions are accurately modeled [127-129]. A recent multiple sequence-alignment based bioinformatics analysis including 1664 receptor-ligand pairs comprising 767 full agonists, 184 partial agonists and 713 antagonists illustrates how large scale data analysis can provide a deeper understanding of the GPCR activation mechanism [118]. This set covered 100 class A GPCRs belonging to 30 subfamilies. Residues at relative positions determined by a multiple sequence alignment were correlated with ligands' physico-chemical descriptors and activities. Such approaches provide information that, in principle, can be exploited in combination with homology models to generate many working hypotheses [155,156]. How can SBDD help in sorting and filtering them to select the relevant ones? In the coming years, drug designers, chemists and biologists will be exposed to a daunting combinatorial explosion of hypotheses, with the cumbersome task to explore them by trial and error [144]. We assume that more integrated multi-software SBDD systems combining state of the art modeling, bioinformatics procedures and bioassay results including site-directed mutagenesis will assist drug designers in these arduous tasks [144,155,157,158].

Concluding remarks

Over the 35 years that have followed the first published work describing SBDD in 1976, computer aided molecular design and SBDD has surmounted several hurdles and has played a key role in the development of several marketed drugs [22,159]. The proof of concept is now well established for SBDD, but cultural shifts are still needed to accept and implement this strategy broadly [159].

Modern genomic investigations and translational research have identified a wealth of new targets. Certainly, in the past decade the number of targets has increased more rapidly than the discovery of new drugs. The plethora of new pharmacological targets must be accompanied by the discovery of new drugs and active classes of ligands in a sustainable manner; otherwise the continuous health care improvement that has characterized the 20th century will fade. The industry is highly profitable thanks largely to revenue from public health schemes and insurance contributions, yet the high costs and risks associated with new drug discovery have been used to justify a growing disengagement from research by the pharmaceutical industry [160]. If not counterbalanced by broad and dedicated investment in SMEs and universities, this process might be highly detrimental to the already damaged social contract between this industry and society in the short and medium term [161].

We believe that homology modeling, docking and virtual screening can be successfully combined to accelerate drug discovery. Just as it was the case for drugs discovered at the beginning of the 20th century, an important part of the reality will always remain hidden, but the synergy of these techniques is part of the 'New Deal' we advocate. With evidence-based SBDD scientists will be put back in control, and they will engage their minds again, taking pride in their "craft".

Acknowledgements

Tina Ritschel thanks the Netherlands Bioinformatics Center (NBIC) and the Deutsche Forschungsgemeinschaft (DFG) for funding. TR is recipient of a personal grant: Ri 2087/1-1 from DFG.

List of abbreviations

ADME/tox	absorption, digestion, metabolism, excretion, and toxicity
ATP	adenosine triphosphate
β 2AR	beta-2 adrenergic receptor
C α	alpha carbon
CoMFA	comparative molecular field analysis
CoMSIA	comparative molecular similarity index analysis
CXCR4	C-X-C chemokine receptor type 4
ECL2	extracellular loop-2
Erk2	Extracellular signal-regulated kinases 2
FBDD	fragment-based drug design
GPCR	G-protein coupled receptor
HIV	human immunodeficiency virus
HTS	high throughput screening
IC50	half maximal inhibitory concentration
ITC	isothermal titration calorimetry
JCAMD	Journal of Computer-Aided Molecular Design
Kd	dissociation constant
MDS	molecular dynamics simulation
MLogP	Moriguchi octanol-water partition coefficient
NR	nuclear receptor
PK	protein kinase
PSA	polar surface area
QSAR	quantitative structure-activity relationship
SBDD	structure-based drug design
RNA	ribonucleic acid
SME	small to medium enterprise
tRNA	transfer RNA
TGT	tRNA-guanine transglycosylase
VLS	virtual ligand screening

Citation

Lounnas V, Ritschel T, Kelder J, McGuire R, Bywater RP, Foloppe N (2013) Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. *Computational and Structural Biotechnology Journal*. 5 (6): e201302011. doi: <http://dx.doi.org/10.5936/csbj.201302011>

References

- Szöllösi-Janze M (2001) *Science in the Third Reich (German Historical Perspectives)*. Oxford, UK: Berg Publishers. 256 p.
- Murray GR (1920) The life history of the first case of myxoedema treated by thyroid extract. *Br Med J* 1920: 359-360.
- Wolf G (2004) The discovery of vitamin D: the contribution of Adolf Windaus. *J Nutr* 134: 1299-1302.
- Hansen SE (2008) The development of adrenal cortical hormones into drugs. *Dan Medicinhist Arbog* 36: 109-132.
- Rosenheim O, King H (1932) The Ring-system of sterols and bile acids. Part II. *J Chem Technol Biotechnol* 51: 954-957.
- Ingle DJ (1950) The biologic properties of cortisone: a review. *J Clin Endocrinol Metab* 10: 1312-1354.
- Stouch TR (2012) The errors of our ways: taking account of error in computer-aided drug design to build confidence intervals for our next 25 years. *J Comput Aided Mol Des* 26: 125-134.
- Marshall GR (2012) Limiting assumptions in structure-based design: binding entropy. *J Comput Aided Mol Des* 26: 3-8.
- Craig A (2003) Astronomers count the stars. Available: <http://news.bbc.co.uk/2/hi/science/nature/3085885.stm>. Accessed 2013 Jan 25.
- Reymond JL, Blum LC, van Deursen R (2011) Exploring the chemical space of known and unknown organic small molecules at www.gdb.unibe.ch. *Chimia (Aarau)* 65: 863-867.
- Gorse AD (2006) Diversity in medicinal chemistry space. *Curr Top Med Chem* 6: 3-18.
- Rubenstein K, Cory C High Throughput Screening: Redefining the Mission. Available: http://pharmalicensing.com/public/articles/view/1005568086_3befc0562a952/. Accessed 2013 Jan 25.
- Rick Mullin (2004) Drug Discovery: As high-throughput screening draws fire, researchers leverage science to put automation into perspective. *Chemical & Engineering News* 82: 23-32.
- Hopkins AL, Groom CR, Alex A (2004) Ligand efficiency: a useful metric for lead selection *Drug Discovery Today* 9: 430-431.
- Vicki Glaser (2009) High Throughput Screening Retools for the Future. Available: http://www.bioitworld.com/BioIT_Article.aspx?id=86836. Accessed 2013 Jan 25.
- Miteva MA (2008) Hierarchical structure-based virtual screening for drug design. *Biotechnol & Biotechnol Equip* 1: 634-638.
- Miteva MA, Olivier Sperandio, Bruno O. Villoutreix (2007) Virtual Ligand Screening for Structure-based Drug Design: Approaches and Progress. *Bioautomation* 7: 104 - 121.
- Foloppe N, Fisher LM, Howes R, Potter A, Robertson AGS, Surgenor AE (2006) Identification of chemically diverse Chk1 inhibitors by receptor-based virtual screening. *Bioorganic & Medicinal Chemistry* 14: 4792-4802.
- Klebe G (2006) Virtual ligand screening: strategies perspective and limitations. *Drug Discovery Today* 11: 580-594.
- Bleicher KH, Böhm HJ, Müller K, Alanine AI (2003) A guide to drug discovery: Hit and lead generation: beyond high-throughput screening. *Nature Reviews Drug Discovery* 2: 369-378.
- Kitchen DB, Decornez H, Furr JR, and Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery* 3: 925-949.
- Seddon G, Lounnas V, McGuire R, van den Bergh T, Bywater RP, et al. (2012) Drug design for ever, from hype to hope. *J Comput Aided Mol Des* 26: 137-150.
- Hartshorn MJ, Murray CW, Cleasby A, Frederickson M, Tickle IJ, et al. (2005) Fragment-Based Lead Discovery Using X-ray Crystallography. *J Med Chem* 48: 403-413.
- Carr R, Hann M (2002) The right road to drug discovery? Fragment-based screening casts doubt on the Lipinski route. *Modern Drug Discov* April, 45-48.
- Davies NG, Browne H, Davis B, Drysdale MJ, Foloppe N, et al. (2012) Targeting conserved water molecules: Design of 4-aryl-5-cyanopyrrolo[2,3-d]pyrimidine Hsp90 inhibitors using fragment-based screening and structure-based optimization. *Bioorg Med Chem* 20: 6770-6789.
- Jeppsson F, Eketjäll S, Janson J, Karlstrom S, Gustavsson S, et al. (2012) Discovery of AZD3839, a potent and selective BACE1

- clinical candidate for the treatment of Alzheimers Disease. *J Biol Chem* Oct 9. [Epub ahead of print].
27. Knehans T, Schüller A, Doan DN, Nacro K, Hill J, et al. (2011) Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *J Comput Aided Mol Des* 25: 263-274.
 28. Erlanson DA (2012) Introduction to fragment-based drug discovery. *Top Curr Chem* 317: 1-32.
 29. Foloppe N (2011) The benefits of constructing leads from fragment hits. *Future Med Chem* (2011) 3: 1111–1115.
 30. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate—a historical perspective. *Drug Discovery Today* 14: 668-675.
 31. Acharya C, Coop A, Polli JE, Mackerell AD Jr (2011) Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr Comput Aided Drug Des* 7: 10-22.
 32. Bursi R, Grootenhuys PD (1999) Comparative molecular field analysis and energy interaction studies of thrombin-inhibitor complexes. *J Comput Aided Mol Des* 13: 221–232.
 33. Kubinyi H (2007) Hydrogen Bonding: The Last Mystery in Drug Design? In: Testa B, van de Waterbeemd H, Folkers G, Guy R, editors. *Pharmacokinetic Optimization in Drug Research: Biological, Physicochemical, and Computational Strategies*. Verlag Helvetica Chimica Acta, Postfach, CH-8042 Zürich, Switzerland. pp. 513-524.
 34. Kraljevic S, Stambrook PJ, Kresimir Pavelic K (2004) Accelerating drug discovery. *EMBO Rep* 5: 837–842.
 35. Vyas V, Jain A, Jain A, Gupta A (2008) Virtual Screening: A Fast Tool for Drug Design. *Sci Pharm* 76: 333–360.
 36. Cheng T, Li Q, Zhou Z, Wang Y, Bryant SH (2012) Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *The AAPS Journal* 14: 133-141.
 37. Ripphausen P, Nisius B, Peltason L, Bajorath J (2010) Quo vadis, virtual screening? A comprehensive survey of prospective applications. *J Med Chem* 53: 8461-8467.
 38. Vogt M, Stumpfe D, Geppert H, Bajorath J (2010) Scaffold hopping using two-dimensional fingerprints: true potential, black magic, or a hopeless endeavor? Guidelines for virtual screening. *J Med Chem* 53: 5707-5715.
 39. Wolf A, Shahid M, Kasam V, Ziegler W, Hofmann-Apitius M (2010) In silico drug discovery approaches on grid computing infrastructures. *Curr Clin Pharmacol* 5: 37-46.
 40. Fukunishi Y (2009) Structure-Based Drug Screening and Ligand-Based Drug Screening with Machine Learning. *Combinatorial Chemistry & High Throughput Screening* 12: 397-408.
 41. Villoutreix BO, Eudes R, Miteva MA (2009) Structure-based virtual ligand screening: recent success stories. *Comb Chem High Throughput Screen* 12: 1000-1016.
 42. Seifert MHJ, Martin L (2008) Essential Factors for Successful Virtual Screening. *Mini Reviews in Medicinal Chemistry* 8: 63-72.
 43. Balakin KV, Kozintsev AV, Kiselyov AS, Savchuk NP (2006) Rational design approaches to chemical libraries for hit identification. *Curr Drug Discov Technol* 3: 49-65.
 44. Wang J (2012) Challenges in Binding Free Energy Calculation Using MM-PB/GBSA. *Bioenerg Open Access* 1:e102. Available: <http://www.omicsgroup.org/journals/2167-7662/2167-7662-1-e102.php?aid=4747>. Accessed 2013 Jan 25.
 45. General IJ, Dragomirova R, Meirovitch H (2011) New Method for Calculating the Absolute Free Energy of Binding: The Effect of a Mobile Loop on the Avidin/Biotin Complex. *J. Phys. Chem. B* 115: 168–175.
 46. Gu RX, Liu LA, Wei DQ, Du JG, Liu L, Liu H (2011) Free Energy Calculations on the Two Drug Binding Sites in the M2 Proton Channel. *J Am Chem Soc* 133: 10817–10825.
 47. Malmstrom RD, Watowich SJ (2011) Using Free Energy of Binding Calculations To Improve the Accuracy of Virtual Screening Predictions *J Chem Inf Model* 51: 1648–1655.
 48. Singh N, Warshel A. (2010) Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. *Protein* 78: 1705-1723.
 49. Fukunishi Y, Mitomo D, Nakamura H (2009) Protein-ligand binding free energy calculation by the smooth reaction path generation (SRPG) method. *J Chem Info Model* 49: 1944-1951.
 50. Jiao D, Golubkov PA, Darden TA, Ren P (2008) Calculation of protein–ligand binding free energy by using a polarizable potential. *PNAS* 105: 6290-6295.
 51. Wang J, Deng Y, Roux B (2006) Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraining Potentials. *Biophysical Journal* 91: 2798-2814.
 52. Duff MR Jr, Grubbs J, Howell EE (2011) Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. *J Vis Exp* 55: e2796.
 53. Chaires JB (2008) Calorimetry and Thermodynamics in Drug Design. *Annu Rev Biophys* 37: 135-151.
 54. Englebienne P, Van Hoonacker Anne, Verhas M (2003) Surface plasmon resonance: principles, methods and applications in biomedical sciences. *Spectroscopy* 17: 255–273.
 55. Ruvinsky AM (2007) Calculations of protein-ligand binding entropy of relative and overall molecular motions. *J Comput Aided Mol Des* 21: 361-370.
 56. Chaplin M Water structure in science. <http://www.lsbu.ac.uk/water/phobic.html>. Accessed 2013 Jan 25.
 57. Head Gordon T (1995) Is water structure around hydrophobic groups clathrate-like? *Proc Natl Acad Sci USA* 92: 8308-8312.
 58. Butler JAV (1937) The energy and entropy of hydration of organic compounds. *Trans Faraday Soc* 33: 229-238.
 59. Lin J, Sahakian DC, de Morais SM, Xu JJ, Polzer RJ, Winter SM (2003) The role of absorption, distribution, metabolism, excretion and toxicity in drug discovery. *Curr Top Med Chem* 3: 1125-1154.
 60. Chodera JD, Mobley DL (2013) Entropy-Enthalpy Compensation: Role and Ramifications in Biomolecular Ligand Recognition and Design. *Annual Review of Biophysics* 42: (Volume publication date June 2013)
 61. Roy S, Bagchi B (2012) Free Energy Barriers for Escape of Water Molecules from Protein Hydration Layer The journal of Physical Chemistry. *J Phys Chem B* 116: 2958-68.
 62. Ritschel T, Philipp C, Kohler PC, Neudert G, Heine A, et al. (2009) How to Replace the Residual Solvation Shell of Polar Active Site Residues to Achieve Nanomolar Inhibition of tRNA-Guanine Transglycosylase. *ChemMedChem* 4: 2012 – 2023.
 63. Kadirvelraj R, Foley BL, Dyekjær JD, Woods RJ (2008) Involvement of Water in Carbohydrate-Protein Binding: Concanavalin A Revisited. *J Am Chem Soc* 130: 16933–16942.
 64. Li Z, Lazaridis T (2007) Water at biomolecular binding interfaces. *Phys Chem Chem Phys* 9: 573-581.
 65. Nguyen CN, Kurtzman Young T, Gilson MK (2012) Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* 137: 044-101.
 66. Ross GA, Morris GM, Biggin PC (2012) Rapid and Accurate Prediction and Scoring of Water Molecules in Protein Binding Sites. *PLoS ONE* 7: e32036.

67. Water...where it matters, when it matters. Available: <http://www.eyesopen.com/SZMAP>. Accessed 2013 Jan 25.
68. A new paradigm in ligand optimization. Available: <http://www.schrodinger.com/WaterMap.php>. Accessed 2013 Jan 25.
69. Genheden S, Luchko T, Gusarov S, Kovalenko A, Ryde U (2010) An MM/3D-RISM Approach for Ligand Binding Affinities. *J Phys Chem* 114: 8505–8516.
70. Genheden S, Mikulskis P, Hu L, Kongsted J, Söderhjelm P, Ryde U (2011) Accurate Predictions of Nonpolar Solvation Free Energies Require Explicit Consideration of Binding-Site Hydration. *J Am Chem Soc* 133: 13081–13092.
71. Ben-Naim A (1990) Solvent effects on protein association and protein folding. *Biopolymers* 29: 567–596.
72. Li Z, Lazaridis T (2006) Thermodynamics of buried water clusters at a protein at ligand binding interface. *J Phys Chem B* 110: 1464–1475.
73. Li Z, Lazaridis T (2005) The effect of water displacement on binding thermodynamics: concanavalin A. *J Phys Chem B* 109: 662–670.
74. Kohler PC, Ritschel T, Schweizer WB, Klebe G, Diederich F (2009) High-Affinity Inhibitors of tRNA-Guanine Transglycosylase Replacing the Function of a Structural Water Cluster. *Chemistry* 15: 10809–10817.
75. Stengl B, Meyer E, Heine A, Brenk R, Diederich F, et al. (2007) Crystal Structures of tRNA-guanine Transglycosylase (TGT) in Complex with Novel and Potent Inhibitors Unravel Pronounced Induced-fit Adaptations and Suggest Dimer Formation Upon Substrate Binding. *J Mol Biol* 370: 492–511.
76. Chakravarty S, Ghersi D, Sanchez R (2011) Systematic assessment of accuracy of comparative model of proteins belonging to different structural fold classes. *J Mol Model* 17:2831–2837.
77. Zhang Y (2008) Progress and challenges in protein structure prediction. *Curr Opin Struct Biol* 18: 342–348.
78. Krieger E, Joo K, Lee J, Lee J, Raman S, et al. (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* 77: 114–122.
79. Venselaar H, Joosten RP, Vrolijk B, Baakman CAB, Hekkelman ML, et al. (2010) Homology modelling and spectroscopy, a never-ending love story. *Eur Biophys J* 39: 551–563.
80. Brooijmans N, Chang YW, Mobilio D, Denny RA, Humblet C (2010) An enriched structural kinase database to enable kinome-wide structure-based analyses and drug discovery. *Protein Sci* 19: 763–774.
81. Knight JDR, Qian B, Baker D, Kothary R (2007) Conservation, Variability and the Modeling of Active Protein Kinases. *PLoS ONE* 2: e982.
82. Protein Data Bank (2011) <http://www.pdb.org>. Accessed 2013 Jan 25.
83. Lundstrom K (2009) An overview on GPCRs and drug discovery: structure-based drug design and structural biology on GPCRs. *Methods Mol Biol* 552: 51–66.
84. Franceschi F, Duffy EM (2006) Structure-based drug design meets the ribosome. *Biochem Pharmacol* 71: 1016–1025.
85. Todd AE, Marsden RL, Thornton JM, Orengo CA (2005) Progress of structural genomics initiatives: an analysis of solved target structures. *J Mol Biol* 348: 1235–1260.
86. Janzen WP, Hull-Ryde A (2011) Kinases as Targets for Drug Discovery in an Academic Setting. *International Drug Discovery*. Available: www.internationaldrugdiscovery.com/images/article/1_Janzen.pdf. Accessed 2013 Jan 25.
87. Chico LK, Van Eldik LJ, Watterson DM (2009) Targeting protein kinases in central nervous system disorders. *Nat Rev Drug Discov* 8: 892–909.
88. Catapano LA, Manji HK (2008) Kinases as drug targets in the treatment of bipolar disorder. *Drug Disc Today* 13: 295–302.
89. Zhang J, Yang PL, Gray NS. (2009) Targeting cancer with small molecule kinase inhibitors. *Nature Rev Cancer* 9: 28–39.
90. Overington J (2010) Kinase inhibitor patents over time. Available: <http://chembl.blogspot.nl/2010/03/kinase-inhibitor-patents-over-time.html>. Accessed 2013 Jan 25.
91. Suhre SJ, Markus Wiederstein M, Sippl MJ (2007) QSCOP—SCOP quantified by structural relationships. *Bioinformatics Application Note*. 23: 513–514.
92. Scheeff ED, Bourne PE (2005) Structural Evolution of the Protein Kinase-Like Superfamily. *PLoS Comput Biol* 5: e49.
93. Hanks SK, Hunter T. (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9: 576–596.
94. Ando T, Imamura H, Suzuki R, Aizaki H, Watanabe T, et al. (2012) Visualization and Measurement of ATP Levels in Living Cells Replicating Hepatitis C Virus Genome RNA. *PLoS Pathog* 3: e1002561.
95. Bywater RP (2012) Protein folding: a problem with multiple solutions. *J Biomol Struct Dyn* Aug 7. [Epub ahead of print]
96. Marcheler-Bauer A, Panchenko AR, Ariel N, Bryant S (2002) Comparison of sequence and structure alignments for protein domains. *Proteins* 48: 439–446.
97. Panchenko AR, Bryant S (2002) A comparison of Position-Specific Score Matrices based on sequence and structure alignments. *Protein Science* 11: 361–370.
98. Toomula N, Kumar DS, Kumar AP (2011) Computational Methods for Protein Structure Prediction and Its Application in Drug Design. *J Proteomics Bioinform* 4: 289–293.
99. Faure G, Aurélie Bornot A, de Brevern AG (2008) Protein contacts, inter-residue interactions and side-chain Modelling. *Biochimie* 90: 626–639.
100. Hartmann C, Antes I, Lengauer T (2007) IRECS: A new algorithm for the selection of most probable ensembles of side-chain conformations in protein models. *Protein Sci* 16: 1294–1307.
101. Kannan N, and Andrew F, Neuwald AF (2004) Evolutionary constraints associated with functional specificity of the CMGC protein kinases MAPK, CDK, GSK, SRPK, DYRK, and CK2 α . *Protein Sci* 13: 2059–2077.
102. Sheinerman FB, Giraud E, Laoui A. (2005) High affinity targets of protein kinase inhibitors have similar residues at the positions energetically important for binding. *J Mol Biol* 352: 1134–1156.
103. Caballero J, Quiliano M, Alzate-Morales JH, Zimic M, Deharo E (2011) Docking and quantitative structure–activity relationship studies for 3-fluoro-4-(pyrrolo[2,1-f][1,2,4]triazin-4-yloxy)aniline, 3-fluoro-4-(1H-pyrrolo[2,3-b]pyridin-4-yloxy)aniline, and 4-(4-amino-2-fluorophenoxy)-2-pyridinylamine derivatives as c-Met kinase inhibitors. *J Comput Aided Mol Des* 25: 349–369.
104. Repasky MP, Murphy RB, Banks JL, GreenwoodJR, Tubert-Brohman I, et al. (2012) Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J Comput Aided Mol Des* 26: 787–799.

105. Spitzer R, Jain AN (2012) Surflex-Dock: Docking benchmarks and real-world application. *J Comput Aided Mol Des* 26: 687–699.
106. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des* 16: 151–166.
107. Kulkarni RG, Srivani P, Achaiah G, Sastry GN (2007) Strategies to design pyrazolyl urea derivatives for p38 kinase inhibition: a molecular modeling study *J Comput Aided Mol Des* 21: 155–166.
108. Assefa H, Kamath S, Buolamwini JK (2003) 3D-QSAR and docking studies on 4-anilinoquinazoline and 4-anilinoquinoline epidermal growth factor receptor (EGFR) tyrosine kinase inhibitors. *J Comput Aided Mol Des* 17: 475–493.
109. Nicklaus MC, Milne GW, Burke TR Jr (1992) QSAR of conformationally flexible molecules: comparative molecular field analysis of protein-tyrosine kinase inhibitors. *J Comput Aided Mol Des* 6: 487–504.
110. Pollack SJ, Beyer KS, Lock C, Müller I, Sheppard D, et al. (2011) A comparative study of fragment screening methods on the p38a kinase: new methods, new insights. *J Comput Aided Mol Des* 25: 677–687.
111. Foloppe N, Fisher LM, Howes R, Kierstan P, Potter A, et al. (2005) Structure-Based Design of Novel Chk1 Inhibitors: Insights into Hydrogen Bonding and Protein-Ligand Affinity. *J Med Chem* 48: 4332–4345.
112. Chema D, Eren D, Yayon A, Goldblum A, Zaliani A (2004) Identifying the binding mode of a molecular scaffold. *J Comput Aided Mol Des* 18: 23–40.
113. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? *Nat Rev Drug Discov* 5: 993–996.
114. Paiva AC, Oliveira L, Horn F, Bywater RP, Vriend G. (2006) Modeling GPCRs. *Ernst Schering Found Symp Proc*. 2: 23–47.
115. Oliveira L, Hulsen T, Lutje Hulsik D, Paiva ACM, Vriend G (2004) Heavier-than-air flying machines are impossible. *FEBS Letters* 564: 269–273.
116. Rasmussen SG, DeVree BT, Zou Y, Kruse AC, Chung KY, et al. (2011) Crystal structure of the β_2 adrenergic receptor-Gs protein complex. *Nature* 477: 549–555.
117. Kobilka BK, Deupi X (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol Sci* 28: 397–406.
118. Wichard JD, ter Laak A, Krause G, Heinrich N, Kühne R, et al. (2011) Chemogenomic Analysis of G-Protein Coupled Receptors and Their Ligands Deciphers Locks and Keys Governing Diverse Aspects of Signalling. *PLoS ONE* 6: e16811.
119. Smith SO (2010) Structure and Activation of the Visual Pigment Rhodopsin. *Annu Rev Biophys* 39: 309–328.
120. Provasi D, Artacho MC, Negri A, Mobarec JC, Filizola M (2011) Ligand-Induced Modulation of the Free-Energy Landscape of G Protein-Coupled Receptors Explored by Adaptive Biasing Techniques. *PLoS Comput Biol* 7: e1002193.
121. Ye S, Zaitseva E, Caltabiano G, Schertler GFX, Sakmar TP, et al. (2010) Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. *Nature* 464: 1386–1389.
122. Scheerer P, Park JH, Hildebrand PW, Kim YJ, Krau N, et al. (2008) Crystal structure of opsin in its G-protein-interacting conformation. *Nature* 455: 497–502.
123. Deupi X, Kobilka B (2007) Activation of G Protein-coupled receptors. *Advances in protein chemist* 74: 137–166.
124. Rubenstein LA, Lanzara RG (1998) Activation of G protein-coupled receptors entails cysteine modulation of agonist binding. *Journal of Molecular Structure (Theochem)* 430: 57–71.
125. Ahuja S, Smith SO (2009) Multiple switches in G protein-coupled receptor activation. *Trends Pharmacol Sci* 30: 494–502.
126. Hofmann KP, Scheerer P, Hildebrand PW, Choe HW, Park JH, et al. (2009) A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem Sci* 34: 540–552.
127. Bywater RP (2005) Location and nature of the residues important for ligand recognition in G Protein-coupled receptors. *J Mol Recognit* 18: 60–72.
128. Oliveira L, Paiva PB, Paiva AC, Vriend G (2003) Identification of functionally conserved residues with the use of entropy-variability plots. *Proteins* 52: 544–552.
129. Sanders MPA, Fleuren WWM, Verhoeven S, van den Beld S, Alkema W, et al. (2011) ss-TEA: Entropy based identification of receptor specific ligand binding residues from a multiple sequence alignment of class A GPCRs *BMC Bioinformatics* 12: 332–341.
130. Rasmussen SGF, Choi HJ, Rosenbaum DM, Kobilka TS, Thian FS, et al. (2007) Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* 450: 383–387.
131. Engel S, Skoumbourdis AP, Childress J, Neumann S, Deschamps JR, et al. (2008) Virtual Screen for Diverse Ligands: Discovery of Selective G Protein-Coupled Receptor Antagonists. *J Am Chem Soc* 130: 5115–5123.
132. Vilar S, Ferino G, Phatak SS, Berk B, Cavasotto CN, Costanzi S (2011) Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. *J Mol Graph Model* 29: 614–623.
133. Wacker D, Fenalti G, Brown MA, Katritch V, Abagayan R, et al. (2010) Conserved binding mode of human β_2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J Am Chem Soc* 132: 11443–11445.
134. Katritch V, Abagyan R (2012) GPCR agonist binding revealed by modeling and crystallography. *Trends in Pharmacological Sciences* 37: 343–352.
135. Kobilka B, Schertler GF (2008) New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol Sci* 29: 79–83.
136. Michino M, Abola E, Brooks CL, Dixon JS, Moul J, et al. (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat Rev Drug Discov* 8: 455–463.
137. Perez-Aguilar JM, Saven JG, Liu R (2012) Human μ Opioid Receptor Models with Evaluation of the Accuracy Using the Crystal Structure of the Murine μ Opioid Receptor. *J Anesth Clin Res* 3:6.
138. Worth CL, Kreuchwig A, Gunnar Kleinau G, Krause G (2011) GPCR-SSFE: A comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics* 12: 185.
139. Topiol S, Sabio M (2009) X-ray structure breakthroughs in the GPCR transmembrane region. *Biochem Pharmacol* 2009 78: 11–20.
140. Horn F, Bettler E, Oliveira L, Campagne F, Cohen FE, Vriend G (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res* 31: 294–297.
141. Costanzi S, Tikhonova IG, Harden TK, Jacobson KA (2009) Ligand and structure-based methodologies for the prediction of the activity of G protein-coupled receptor ligands. *J Comput Aided Mol Des* 23: 747–754.
142. Foloppe N, Benwell K, Brooks TD, Kennett G, Knight AR, et al. (2009) Discovery and functional evaluation of diverse novel human CB1 receptor ligands. *Bioorganic & Medicinal Chemistry Letters* 19: 4183–4190.

143. Poulsen A, Bjørnholm B, Gundertofte K, Pogozheva ID, Liljefors T (2003) Pharmacophore and receptor models for neurokinin receptors. *J Comput Aided Mol Des* 11: 765-783.
144. Launay G, Téletchéa S, Wade F, Pajot-Augy E, Gibrat JF, Sanz G (2012) Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng Des Sel* 25: 377-386.
145. Jain V, Saravanan P, Arvind A, Mohan CG (2011) First pharmacophore model of CCR3 receptor antagonists and its homology model-assisted, stepwise virtual screening. *Chem Biol Drug Des* 77: 373-387.
146. Tebben AJ, Schnur DM (2011) Beyond rhodopsin: G protein-coupled receptor structure and modeling incorporating the beta2-adrenergic and adenosine A(2A) crystal structures. *Methods Mol Biol* 672: 359-386.
147. Davies MN, Bayry J, Tchilian EZ, Vani J, Shaila MS, et al. (2009) Toward the Discovery of Vaccine Adjuvants: Coupling *In Silico* Screening and *In Vitro* Analysis of Antagonist Binding to Human and Mouse CCR4 Receptors. *PLoS ONE* 4: e8084.
148. 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.
149. Radestock S, Weil T, Renner S (2008) Homology Model-Based Virtual Screening for GPCR Ligands Using Docking and Target-Biased Scoring. *J Chem Inf Model* 48: 1104-1117.
150. Chen JZ, Wang J, Xie XQ (2007) GPCR structure-based virtual screening approach for CB2 antagonist search. *J Chem Inf Model* 2007 47: 1626-1637.
151. Schlegel B, Laggner C, Meier R, Langer T, Schnell D, et al. (2007) Generation of a homology model of the human histamine H3 receptor for ligand docking and pharmacophore-based screening. *J Comput Aided Mol Des* 21: 437-453.
152. Singh N, Chevé G, Ferguson DM, McCurdy CR (2006) A combined ligand-based and target-based drug design approach for G-protein coupled receptors: application to salvinorin A, a selective kappa opioid receptor agonist. *J Comput Aided Mol Des* 20: 471-493.
153. Bissantz C, Bernard P, Hibert M, Rognan D. (2003) Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* 50: 5-25.
154. Bröer BM, Gurrathet M, Höltje HD (2003) Molecular modeling studies on the ORL1-receptor and ORL1-agonists *J Comput Aided Mol Des* 17: 739-754.
155. Sanders MP, Verhoeven S, De Graaf C, Roumen L, Vrooling B, et al. (2011) Snooker: a structure-based pharmacophore generation tool applied to class A GPCRs. *J Chem Inf Model* 51: 2277-2292.
156. Sanders MPA, Roumen L, van der Horst E, Lane JR, Vischer HF, et al. (2012) A Prospective Cross-Screening Study on G-Protein-Coupled Receptors: Lessons Learned in Virtual Compound Library Design. *J Med Chem* 55: 5311-5325.
157. Levit A, Barak D, Behrens M, Meyerhof W, Niv MY (2012) Homology Model-Assisted Elucidation of Binding Sites in GPCRs. *Methods Mol Biol*. 914:179-205.
158. Roumen L, Sanders MPA, Bas Vrooling B, de Esch IJP, Jacob de Vlieg, et al. (2011) *In Silico Veritas*. The Pitfalls and Challenges of Predicting GPCR-Ligand Interactions. *Pharmaceuticals* 4: 1196-1215.
159. Beddell CR, Goodford PJ, Norrington FE, Wilkinson S, Wootton R (1976) Compounds designed to fit a site of known structure in human haemoglobin. *Br J Pharmacol* 57: 201-209.
160. Silverman E (2012) Sanofi CEO: Who needs Big Pharma Scientists? Available: <http://www.pharmalot.com/2012/03/sanofi-ceo-who-needs-big-pharma-scientists/>. Accessed 2013 Jan 25.
161. Santoro MA (2006) Charting a Sustainable Path for the Twenty-First Century Pharmaceutical Industry. In: Santoro MA and Gorrie TM, editors. *Ethics and the Pharmaceutical Industry*. Cambridge University Press. pp. 1-9.

Keywords:

Rational drug design, virtual screening, protein kinase, G-protein coupled receptors, ligand binding thermodynamics

Competing Interests:

The authors have declared that no competing interests exist.



© 2013 Lounnas et al.

Licensee: Computational and Structural Biotechnology Journal.

This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are properly cited.

What is the advantage to you of publishing in *Computational and Structural Biotechnology Journal (CSBJ)* ?

- ✚ Easy 5 step online submission system & online manuscript tracking
- ✚ Fastest turnaround time with thorough peer review
- ✚ Inclusion in scholarly databases
- ✚ Low Article Processing Charges
- ✚ Author Copyright
- ✚ Open access, available to anyone in the world to download for free

WWW.CSBJ.ORG