Computational methods used in Hit-to-Lead and Lead Optimization stages of structure-based drug discovery

Alexander Heifetz[1,2*], Michelle Southey[1], Inaki Morao[1], Andrea Townsend-Nicholson[2] and Mike J. Bodkin[1]

[1]Evotec Ltd, 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire, OX14 4RZ, UK
[2]Institute of Structural & Molecular Biology, Research Department of Structural & Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom

*Corresponding author

Keywords: Structure based drug design; molecular dynamics; simulation; Hit-to-lead, Lead optimization; G-protein coupled receptor; docking;

Summary

GPCR modeling approaches are widely used in the hit-to-lead (H2L) and lead optimization (LO) stages of drug discovery. The aims of these modeling approaches are to predict the 3D structures of the receptor-ligand complexes, to explore the key interactions between receptor and ligand and to utilize these insights in the design of new molecules with improved binding, selectivity or other pharmacological properties. In this book chapter, we present a brief survey of key computational approaches integrated with hierarchical GPCR modeling protocol (HGMP) used in hit-to-lead (H2L) and in lead optimization (LO) stages of structure based drug discovery (SBDD). We outline the differences in modeling strategies used in H2L and LO of SBDD. We illustrate how these tools have been applied in three drug discovery projects.

1. Introduction

1.1 GPCRs are cell surface receptors that contain seven transmembrane helices and constitute the largest superfamily of membrane proteins, regulating almost every aspect of cellular activity [1]. GPCRs have enormous physiological and biomedical importance, being the primary site of action of 40% of all prescribed drugs today [2]. There are over 800 human GPCRs known today [3, 4], involved in a diversity of diseases, including cancer, pain, inflammation, depression, anxiety [5]. Despite this, drugs have been developed just for 50 of these GPCRs. This renders GPCRs as one of the most important classes of current pharmacological targets [3, 5].

1.2 Recent advances in X-ray crystallography of GPCR experiencing its ‘renaissance’ [2, 6-10], however, crystal structures are still not currently feasible for every receptor or receptor-ligand complex [11]. This significantly limits the ability of the crystallography to guide SBDD for GPCR targets in “real-time” [11]. Furthermore, experimentally

determined structures represent just a few snapshots of what we know are very dynamic receptors and as a consequence offer only limited insights into the overall conformational space and related function of GPCR [11].

1.3 In the absence of crystallographic data, GPCR modeling is often the only practical alternative to guide SBDD [1, 12-14]. Modern computational approaches can address such key issues such as GPCR flexibility [15] and ligand-induced dynamics, ligand kinetics (k_{on}/k_{off} rates) [16-19], prediction of water positions [20] and their role in ligand binding and prediction of the effects of mutations on ligand binding. However, the ultimate goals of any GPCR modeling protocol are: 1) to predict the structures of the complexes between ligands and the target receptor, 2) to explore the key interactions between the ligand, surrounding residues and water molecules and 3) to utilize these insights in the design of the next generation of the lead compounds with improved binding, selectivity or other pharmacological properties. The success of any GPCR modeling protocol applied in SBDD is always measured by decreased time and cost of the synthetic effort [14, 21].

1.4 Hit to lead (H2L) [22] is defined as early stage of drug discovery also known as lead generation (Figure 1A). In H2L small molecule hits from a high throughput screen (HTS) or from virtual screening (VS) are evaluated and undergo limited optimization to identify promising lead compounds as illustrated in Figure 1A. Through the limited H2L optimization steps, the affinities of these primary hits are often improved by several orders of magnitude to the nanomolar (10^{-9} M) range [22]. To achieve improvement in affinity it is usually sufficient to modify the hit in such way that it will generate additional interaction/s with the target receptor compare to the primary hit.

<Figure 1A, here>

1.5 Lead optimization (LO) [21] phase of drug discovery (Figure 1B) is usually defined as the process of bringing a chemical series to clinical trials through iterative steps of design and testing. Compared to H2L, the initial lead compound(s) in LO often already demonstrated significant potency against the target. However, the affinity, selectivity or other pharmacological properties might need further optimization. The key challenge in LO is to improve of what are often already potent compounds. This requires detailed information on the interactions between the ligand and its corresponding target and off-target receptors. Any modeling input must therefore be accurate and give reliable insights at the molecular level.

<Figure 1B, here>

1.6 Integrating of GPCR homology modelling with other modelling approaches such as docking and fragment molecular orbitals (FMO) can be powerful tool to guide SBDD [21, 23], as it provides an accurate and comprehensive list of strong, weak, or repulsive interactions between the ligand and its surrounding residues. Such information is highly useful in rational design of the next generation of lead compounds in terms of modifications, scaffold replacement (scaffold hoping), linking (specifically in case of fragment-based drug discovery) or extension of chemical moieties to form stronger or new interactions with the protein or alternatively to remove repulsions. It can also be helpful in analysis of the ligand-water-protein network, to distinguish between energetically favorable and unfavorable water molecules and to design ligands that can interact or displace certain waters. FMO energy terms can be efficiently used as descriptors in QSAR modelling to predict the binding affinities of new molecules [24].

1.7 In this book chapter, we will describe one of many GPCR modelling protocols named 'hierarchical GPCR modelling protocol (HGMP [21, 25, 26], Figure 2). HGMP has been developed by Evotec Ltd and University of Oxford to support SBDD programs. HGMP generates a 3D model of GPCR structures and its complexes with small molecules by applying computational methods. In 'Methods' section we will describe how HGMP is integrated with other SBDD tools like: docking, molecular dynamic simulations, FMO, water molecules predictors and KNIME. In 'Notes' section we will illustrate how these tools were use in 3 H2L and LO projects.

2. METHODS

2.1 Constructing GPCR models

2.1.1 Traditional GPCR homology modeling approaches [13] often involve the following steps: (1) sequence alignment between the modeled receptor and an appropriate template, (2) homology modeling and model refinement and (3) docking of ligands into the binding site. The key cons of such 'static' approaches is that the modeled receptor is practically a 'copy' of the original template and therefore some of the critical structural features are often lost. This significantly reduces the relevance of such models and their ability to guide SBDD. This is particularly problematic in the LO when information on the fine details of the system is highly important.

2.1.2 Modern (dynamic) GPCR modeling protocols [13, 27] have moved beyond the use of static homology modeling approaches by performing the type of extensive refinement and exploration of both structure and flexibility that is required to drive SBDD. To address the various challenges of GPCR drug discovery programs, these contemporary approaches are encapsulated as

toolkits that can be flexibly assembled into workflows tailored to the specific needs of each project. The ability to incorporate experimental data during the modeling is another important factor that can enhance the effectiveness of these workflows. An example of such a workflow is the hierarchical GPCR modeling protocol (HGMP - Figure 2).

<Figure 2, here>

2.1.3 Hierarchical GPCR modeling protocol (HGMP) [25] (Figure 2) - generates a GPCR model and its potential complexes with small molecules by applying a series of computational methods incorporated mainly in molecular operating system (MOE, Chemical Computing Group, version 2016.08). The protocol makes use of homology modeling followed by MD simulations and docking (flexible docking if required) to predict binding poses and functions of ligands. The HGMP is practically a toolbox for GPCR modeling that can be 'tailored' for project needs where experimental data can be easily fed in. It is equipped with GPCR-specific "plugins", including a GPCR-likeness assessment score (GLAS) to evaluate model quality and a pairwise protein comparison method (ProS) used to cluster structural data and distinguish between different activation substates. The HGMP has been applied in a number of industrial drug design projects, which have also led to further refinements of the protocol (see Notes 3.1, 3.2 and 3.3). Even in cases where the sequence identity to the target is very low, careful model building in conjunction with site-directed mutagenesis and binding assays can be very useful in aiding the future direction of a drug discovery program or indeed rationalizing

2.2 Generating of the GPCR-ligand complex

2.2.1 Having the model of the receptor in hand, the next step is often predicting of the receptor-ligand complex, this process is called molecular docking. Predicting this complex is highly important if we want to study the interactions between the ligand and receptor and to guide the SBDD. As numerous docking approaches have been reviewed in the literature [28] quite recently we here survey briefly the unique challenges and docking protocols relevant to GPCRs.

2.2.2 **Docking protocols** [28, 29] are the molecular modelling processes aimed to explore the interaction between ligand and protein. The ultimate goal of any docking protocol is to predict the bioactive conformation of the ligand and its place and orientation inside of the receptor binding site named as 'docking pose' or 'binding mode'. The docking procedure consists of two sequential tasks. Firstly, flexible placement of the ligand in a pre-defined binding site of the receptor and then scoring the poses of the docked ligands. Both posing and scoring phases are equally important and can be carried out by very different methodologies depending on how exhaustive the conformational sampling of both ligand and protein are considered.

2.2.3 Some commercial available **docking suites of programs** are AutoDock [30], AutoDock Vina [31], MOE [21], FlexX [32], GOLD [33] and Glide [34]. Different search algorithms are designed to predict the bioactive conformation of the studied compounds through the evaluation of the interactions between ligands and targets [29]. An increase in the quality of the ligand docking can be gained by consideration of flexibility of the modelled system.

2.2.4 **Scoring and re-ranking**: In many of our projects (see **Note** 3.4), we used AMBER interaction energy to rescore and re-rank docking poses. We used the MM_PBSA/GBSA approach [35] to calculate the AMBER interaction energy.

[36] This approach, while subject to the same limitations of all force field based methods, was able to accurately predict relative binding affinities between ligand and protein and was therefore selected as a reliable method to rescore and to rank docking poses [37].

2.2.5 Flexible docking - typical docking protocols keep the receptor (largely) rigid, and so cannot address the issue of receptor flexibility. As these protocols do not take into account the ligand-induced (or ligand-stabilized) conformation of the receptor, it makes it harder to rationalize the effects of ligand binding in terms of activation or deactivation (agonists and antagonists, respectively). Some docking approaches like induced fit docking (IFD) introduced in Autodock 4 [38], AutoDock Vina and Schrödinger assign limited flexibility to the sidechains of key residues. However, this approach is slightly artificial and is an unsatisfactory solution to the general problem of receptor flexibility. The ensemble docking protocol, implemented in GOLD [33], performs docking into multiple states of the same receptor but it is highly governed by the availability of the structural information on the targeted receptor. The perfect scenario would be if the bioactive conformation of the docked ligand was known prior to the docking simulation.

<Figure 3, here>

2.2.6 HGMP-C4XD integrates HGMP with experimental NMR based technology (C4XD) (Figure 3). The C4XD [39] was developed by C4X Discovery Ltd to explore how molecules behave in physiologically relevant solution. C4XD demonstrated that small molecules exist in relatively few conformations in solution and that one of those conformations closely resembles the bioactive form – but which one? Next, during the docking we limit the ligand conformation

space only to the most populated conformations found by C4XD and assign the flexibility to the receptor. The combination of HGMP and the C4XD approaches allows the isolation of the bioactive conformation of the ligand and the identification of the key pharmacophoric features required for GPCR-ligand binding and selectivity. These structural insights are essential for the refinement of GPCR models, for addressing the ligand induced receptor flexibility and for the rationalization of the ligand binding.

2.2.7 An additional way to place the ligand inside of the receptor is to overlay it on top of an already bound ligand (template) usually extracted from the crystal structure. The most common software to perform molecular overlays is ROCS, from OpenEye [40]. An additional minimization of the ligand within the active site is needed after in order to remove clashes with the receptor.

2.2.8 **ROCS protocol** [40] is the most common shape-based superposition method employed in industry nowadays. ROCS performs shape-based overlays of conformers of candidate molecules to a query molecule (template) in one or more conformations. The overlays can be performed very quickly because the molecules can be described as atom-centered Gaussian functions. ROCS maximizes the rigid overlap of these Gaussian functions and thereby maximizes the shared volume between a template and a single conformation of a database molecule. ROCS is therefore used in ligand-based drug design in the absence of the target structure. Despite its simplicity it has shown a similar performance and consistency to other structure-based approaches in virtual screening. Moreover, ROCS has also been incorporated into docking workflows where the obtained ROCS overlay is used as initial placement/pose

within the active site and has also been embedded in alignment-dependent 3D QSAR analyses.

2.3 Exploration of the dynamic nature of GPCRs

2.3.1 GPCRs are, by functional necessity, very dynamic entities. Molecular dynamic (MD) simulation therefore provides an important source of structural and functional information for these receptors (as described in detail in chapter 6 of this book) [15]. MD can be used in a variety of ways including refinement of the homology model in a more realistic membrane environment, exploration of ligand-induced flexibility and function, the analysis of solvent, the effect of mutation on receptor stability and exploration of ligand binding and dissociation kinetics [41, 42]. MD trajectories are often used to generate an ensemble of possible receptor substates. The ProS and GLAS methods outlined in 2.1.3 were developed to explore the structural data generated within MD simulations and to help distinguish between different GPCR substates.

2.3.2 MD simulations also allow one to explore the possibility of allosteric and cryptic binding pockets. Cryptic binding pockets are not exposed to bulk solvent all of the time and so may be hidden in certain crystallographic structures. MD allows these sites to manifest themselves, enabling docking and similar protocols to be followed in the usual manner. Simulations are as well essential for the understanding of allosteric modulation [43, 44]. In some cases, however, full MD simulation may not be required, for example when just local refinement of a homology model is required. In these cases “low-mode” molecular dynamics (LowModeMD) simulation can provide a more rapid solution [45]. LowModeMD, as implement in MOE (Chemical Computing Group), is based on perturbing an existing conformation along a trajectory using initial atomic

velocities with kinetic energy concentrated on the low-frequency vibrational modes, followed by energy minimization.

2.3.3 Residence time and MD - It has been recently demonstrated that GPCR modeling and MD simulation can be a promising tool for the exploration and structural rationalization of ligand-receptor residence time (RT) [15, 16, 18, 46, 47]. The definition of the RT is the length of time for which a small molecule stays bound to its receptor target [48]. The current challenge is the timescale: the millisecond timescales of conventional MD are incompatible with the typical RTs of drugs (up to hours) [15, 46]. To overcome this encounter new approaches to extend MD timescales have been developed. These include: (1) Markov State Models (MSM) - a very powerful method to describe dynamical processes between defined states in MD simulations [14] (2) Metadynamics-based approaches that employ MSM to calculate off-rates based on the transitions between the intermediate (calculated) and predefined end states, and (3) Scaled MD - another approximate approach to rank ligands by their off-rates [46, 47].

2.4 Exploring receptor-ligand interactions

2.4.1 The understanding of binding interactions between a protein and a small molecule plays a key role in the rationalization of potency, selectivity and kinetics. However, even with the crystal structure in hand, visual inspection and force-field based molecular mechanics calculations cannot always explain the full complexity of the molecular interactions that are so critical in LO. Quantum mechanical methods have the potential to address this shortcoming, but the high computational cost has typically made the use of these calculations impractical.

2.4.2 **Fragment Molecular Orbital (FMO) method** [24] (Figure 4A) is widely used by us for protein-ligand binding calculations and drug design because it offers substantial computational savings over traditional QM methods [24, 49]. By dividing the system, both ligand and receptor, into smaller pieces and performing QM calculations on these fragments, one can achieve high efficiency. A typical FMO calculation on a GPCR-ligand complex takes approximately 4h on 36 CPU cores to complete, which is significantly faster than the equivalent classical QM calculations. Recently, we have demonstrated that FMO can be even faster (secs instead of hours) without compromising the accuracy by combining it with density-functional tight-binding (DFTB) method [50].

<Figure 4A and 4B, here>

2.4.3 Using FMO, one can take any protein-ligand complex and calculate a list of interactions and their chemical natures. Many of these interactions are difficult to detect or quantify with non-QM methods [49]. This information is very useful in guiding rational LO in terms of ligand modifications such as scaffold replacement and linking or the extension of chemical moieties to form stronger or new interactions with the protein [51].

2.5 Predicting role of water molecules in receptor-ligand binding

2.5.1 It is known that water-mediated interactions between ligands and receptor are extremely common and highly significant for binding and kinetics [17, 44]. Yet only high-resolution crystal structures are able to give any reliable indication as to the presence of water molecules. Displacement of these key water

molecules can directly affect the ligand binding affinity and it is in the scope of SBDD programs to design compounds that can interact with or efficiently displace these water molecules. The prediction of water molecule networks and their perturbation is also critical in terms of its relationship to kinetics and residence time (see chapter 9 of this book), as has been demonstrated for a series of adenosine A2A receptor antagonists [17].

2.5.2 Several methods (WaterMap [52], WaterFLAP [20], WaterDock [53], AutoDock Vina and 3D-RISM [54]) enable a relatively rapid prediction of water molecule sites and estimation of the energy penalty for water displacement. They can help medicinal chemists to decide whether to interact with or displace a certain water molecule, if a particular sub-pocket of the receptor can be explored by hydrophobic moieties or if a displaced water has to be substituted by a group that mimics the hydrogen bond network. These methods are suitable for both H2L and LO.

2.5.3 Most of these methods are based on MD or Monte Carlo (MC) simulations and observing the peaks in water density can provide the location of water binding sites [55, 56]. However, these calculations can be time-consuming to run, especially with buried cavities, due to the long time it takes for water to permeate within the protein. Grand canonical MC methods [57] can significantly reduce the length of the simulation. This has led to a number of attempts to develop faster methods. JAWS for example is a grid-based MC method that estimates the free energy of displacing a water molecule into bulk. An integral theory approach (3D-RISM [54]) has also reported success in predicting solvation structure within ligand-binding sites and protein cavities. Short molecular simulations can be used as the data for inhomogeneous fluid solvation theory (IFST). This method has the distinct advantage that the free

energy can be broken down into enthalpic and entropic components. IFST also forms the framework for WaterMap [52].

2.6 Combining individual tools into integrated workflow engines

2.6.1 GPCR modeling and SBDD is a multitask process comprised of sequential steps (Figure 5). There is a desire to automate and standardize this process and make it more user-friendly so that less experience users can also work with it.

2.6.2 **Pipeline-Pilot and KNIME** [58, 59] are the most commonly used software packages (commercial and open source, respectively) that automate the modeling process and enable an easy concatenation of the individual tools (nodes) into an integrated workflow. Given the extensive interest in creating new therapeutics based on novel GPCR targets, modeling methodologies that are as streamlined, rapid, precise and accurate as possible are highly desirable and it is expected that an increasing number of workflows will become available in future.

< Figure 5 here>

3. Notes

3.1 In the absence of the structural information of the receptor target, the design of new compounds in a medicinal chemistry programs typically relies purely on SAR data. However, interpreting such data in isolation from specific knowledge of the protein can be challenging and even misleading [14]. Therefore, any additional means that

can build confidence in the SAR interpretation and generate novel structure-based hypotheses is potentially very useful. As a result, GPCR modeling is used to bridge the gap and facilitate SBDD. The introduction of experimental data like SAR into a modelling process allows a refinement of the GPCR models to a degree that is not possible with homology modelling alone and provides a deeper rationalization of ligand binding and selectivity. In this way, modelling methods should be designed to accommodate experimental data in their algorithms and be flexible enough to deal with the wide variety of challenges that drug discovery programs face.

3.2 HGMP can take advantage of the experimental data that can be fed into the modeling process to add extra accuracy and confidence in the modeling outcomes. The use of the HGMP in 'real' drug discovery projects is demonstrated below.

3.3 ***Fighting obesity with a sugar-based library*** [60] - Obesity is an increasingly common condition. Antagonism of the melanin-concentrating hormone-1 receptor (MCH-1R) has been widely reported as a promising therapeutic avenue for obesity treatment. However, discovery and optimization of new compounds targeting MCH-1R has been hindered by a lack of structural information about the MCH-1R and low high throughput screening (HTS) success rates. In this H2L project, we combined HGMP (see **Methods** 2.1.3) with the screening of a diverse library of sugar-based compounds from the VAST technology (Versatile Assembly on Stable Templates [60]). The GPCR-VAST method provides a good example of how ligand SAR data, when combined with modeling, can provide a useful source of structural information on GPCR binding sites and for SBDD.

<Figure 6 here>

3.3.1 The 490 VAST compounds obtained from this library were screened against MCH-1R, resulting in the discovery of a moderately potent MCH-1R antagonist, ACL21823 (IC_{50} = 306 nM, see Figure 6). The discovery of ACL21823 was utilized in the construction of a MCH-1R model and in the refinement of its binding site. We used HGMP (see **Methods** 2.1.3) to model the MCH-1R and the flexible docking protocol of GOLD (see **Methods** 2.2.5) to dock the VAST hits into MCH-1R receptor model. The scoring and re-ranking was performed with AMBER interaction energy (see **Methods** 2.2.4). The usefulness of this method in H2L was demonstrated by a structure-based VS, which achieved a hit rate of 14% and yielded 10 new chemotypes of MCH-1R antagonists including EOAI3367472 (IC_{50} = 131 nM) and EOAI3367474 (IC_{50} = 213 nM).

3.4 *Discovery of selective 5-HT_{2C} agonists for the treatment of metabolic disorders*

[61] - In this LO project, which was performed prior to the publication of the 5-HT_{2B} and 5-HT_{1B} crystal structures, the challenge was to optimize 5-HT_{2c} binders and convert them into strong agonists that were unable to activate 5-HT_{2A} and 5-HT_{2B} receptors. It is known that for effective antagonism, it is sufficient for ligands just to occupy a relevant receptor site in order to inhibit the binding of endogenous ligands. However, agonist discovery has the additional complication and requirement that the ligand must not only be able to both occupy the receptor site but also be able to activate the receptor. Agonist binding should elicit conformational changes in the receptor that result in activation of intracellular G-proteins and/or β-arrestins which, in turn, can modulate the activity of downstream effectors within the cell. The mechanism and structural changes associated with the activation of GPCRs remains unclear, making agonist design quite challenging.

3.4.1 To explore 5-HT_{2C} activation mechanism and to design compounds that would promote receptor activation, HGMP was applied (see **Methods** 2.1.3) to model

both the active and inactive receptor conformations, referred to as 5-HT_{2C}^{active} and 5-$HT_{2C}^{inactive}$, respectively. Ensemble docking with GOLD (see **Methods** 2.2.5) was used to predict the binding modes of lead compounds in 5-HT_{2A}, 5-HT_{2B} and 5-HT_{2C}. It was proposed that agonists enter deeply into 5-HT_{2C} binding site and interact simultaneously with both TM3 and TM6, thus increasing the overall stability of 5-HT_{2C}^{active} and promoting activation. In parallel, we modeled off-targets 5-HT_{2A} and 5-HT_{2B} to filter out compounds from the 5-HT_{2C}^{active} screen that might also bind to these two receptors. We also employed our hERG modeling [62] to take into account the hERG liability of our lead compounds. The final outcome was the discovery of a novel compound **10** (EC_{50} = 8.4 / 762 / 73 nM for 5-$HT_{2C\,/\,2A\,/\,2B}$ and hERG inhibition of 11% at 10μM) [61].

3.5 ***Case study 3:*** **Discovery of potent & selective OX_2 receptor antagonists** [63] - The orexin receptors (OX_1 and OX_2) are linked to a range of different physiological functions including the control of feeding, energy metabolism, modulation of neuro-endocrine function and regulation of the sleep-wake cycle. The key challenges of this project were to increase the OX_2 activity and selectivity of lead compounds over OX_1. This was particularly difficult as OX_1 and OX_2 receptors share over 80% sequence identity at the amino acid level. This project was completed before the crystal structures of OX_1 and OX_2 were released.

3.5.1 HGMP was applied (see **Methods** 2.1.3) to model both OX_1 and OX_2 receptors. We used MD simulation (see **Methods** 2.3.1) as implemented in GROMACS [21, 64] to explore OX_1/ OX_2 selectivity. MD suggested that differences in intra-helical interactions resulted in differences in TM conformation and in the topology of the binding pocket. The differences identified were small but sufficient to design molecules with OX_2 selectivity. This rational design

significantly decreased the amount of synthesis required by focusing effort on the relevant portion of the ligand structure, as outlined in Figure 7. The final compound, EP-009-0513, had K_i values of 4,363 and 5.7 nM for OX_1 and OX_2, respectively.

<Figure 7 here>

3.6 Conclusion - Modern GPCR modeling protocols [65], such as the HGMP, have gone beyond the use of static models to allow for the type of detailed exploration of GPCR-ligand structures required to drive H2L and LO. These methods permit the prediction of GPCR substates in a way that is not possible with static homology modeling alone. The practicality and efficiency of GPCR modeling integrated with other modeling tools is enhanced by experimental data and by the availability of structural information on the targeted GPCR, satisfying the immediate need of the drug discovery process for the information needed to drive SBDD effectively.

3.7 Future challenges - Despite a huge effort by the pharmaceutical industry to design novel drugs for GPCR targets, there is tremendous attrition along R&D pipelines [48]. Many promising drug candidates eventually fail in clinical trials due to a demonstrated lack of efficacy. A retrospective analysis of those that have successfully made it to the market has revealed that their beneficial effects in patients may be attributed to their long drug-target residence times (RTs) - the length of time for which a drug (ligand) stays bound to its receptor target [48]. There is substantial evidence that ~70% of long RT therapeutics displayed higher efficacy than comparable faster-dissociating drugs, supporting a growing recognition that drug-target RT may be of even greater importance than affinity, therapeutically [66].

3.8 Recently several notable reviews [48, 66, 67] have emphasized the crucial role of RT optimisation in the early phases of drug discovery, suggesting that detailed structure-based studies of RT should be introduced in the early phases of drug discovery to prevent “fail late, fail expensive” scenarios. Efforts to include RT in the drug development process have focused on the adoption of either experimental or computational approaches (see **Methods** 2.3.3). Although each approach is very promising they only provide half of the whole picture. Experimental methods can measure the RT but cannot rationalize why certain compounds have longer RTs than the others or suggest ways to modify the structure of the ligand to improve its RT profile. On the other hand, computational methods are only able to provide this essential information if robust experimental data are available. Combining experimental and computational tools, as described in chapter 15 of this book, is a highly encouraging step towards addressing the RT in early stages of H2L and LO.

3.9 Experience has shown that significant progress in technology R&D and ‘know-how’ for GPCR SBDD can only be achieved when there are good interdisciplinary collaborations between experimental and theoretical groups [1].

ACKNOWLEDGMENT

A.H. and A.T.-N. would like to acknowledge the support of EU H2020 CompBioMed project (http://www.compbiomed.eu/, 675451) and the BBSRC Flexible Interchanger Programme project (BB/P004245/1).

4.0 Reference

1. Heifetz A, Schertler GF, Seifert R, Tate CG, Sexton PM, Gurevich VV, Fourmy D, Cherezov V, Marshall FH, Storer RI, Moraes I, Tikhonova IG, Tautermann CS, Hunt P, Ceska T, Hodgson S, Bodkin MJ, Singh S, Law RJ, Biggin PC, (2015), GPCR structure, function, drug discovery and crystallography: report from Academia-Industry International Conference (UK Royal Society) Chicheley Hall, 1-2 September 2014, Naunyn Schmiedebergs Arch Pharmacol. 388: 883-903.
2. Shonberg J, Kling RC, Gmeiner P, Lober S, (2015), GPCR crystal structures: Medicinal chemistry in the pocket, Bioorg Med Chem. 23: 3880-3906.
3. Wise A, Gearing K, Rees S, (2002), Target validation of G-protein coupled receptors, Drug Discov Today. 7: 235-246.
4. Rask-Andersen M, Masuram S, Schioth HB, (2014), The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication, Annu Rev Pharmacol Toxicol. 54: 9-26.
5. Dohlman HG, (2015), Thematic Minireview Series: New Directions in G Protein-coupled Receptor Pharmacology, J Biol Chem. 290: 19469-19470.
6. Jazayeri A, Andrews SP, Marshall FH, (2017), Structurally Enabled Discovery of Adenosine A2A Receptor Antagonists, Chem Rev. 117: 21-37.
7. Jazayeri A, Dias JM, Marshall FH, (2015), From G Protein-coupled Receptor Structure Resolution to Rational Drug Design, J Biol Chem. 290: 19489-19495.
8. Cooke RM, Brown AJ, Marshall FH, Mason JS, (2015), Structures of G protein-coupled receptors reveal new opportunities for drug discovery, Drug Discov Today. 20: 1355-1364.
9. Congreve M, Dias JM, Marshall FH, (2014), Structure-based drug design for G protein-coupled receptors, Prog Med Chem. 53: 1-63.
10. Topiol S, Sabio M, (2009), X-ray structure breakthroughs in the GPCR transmembrane region, Biochem Pharmacol. 78: 11-20.
11. Topiol S, (2013), X-ray structural information of GPCRs in drug design: what are the limitations and where do we go?, Expert Opin Drug Discov. 8: 607-620.
12. Topiol S, Sabio M, (2015), The role of experimental and computational structural approaches in 7TM drug discovery, Expert Opin Drug Discov. 10: 1071-1084.
13. Tautermann CS, Gloriam DE, (2016), Editorial overview: New technologies: GPCR drug design and function-exploiting the current (of) structures, Curr Opin Pharmacol. 30: vii-x.
14. Biggin PC, Aldeghi M, Bodkin MJ, Heifetz A, (2016), Beyond Membrane Protein Structure: Drug Discovery, Dynamics and Difficulties, Adv Exp Med Biol. 922: 161-181.
15. Tautermann CS, Seeliger D, Kriegl JM, (2015), What can we learn from molecular dynamics simulations for GPCR drug design?, Comput Struct Biotechnol J. 13: 111-121.
16. Latorraca NR, Venkatakrishnan AJ, Dror RO, (2017), GPCR Dynamics: Structures in Motion, Chem Rev. 117: 139-155.
17. Guo D, Pan AC, Dror RO, Mocking T, Liu R, Heitman LH, Shaw DE, AP IJ, (2016), Molecular Basis of Ligand Dissociation from the Adenosine A2A Receptor, Mol Pharmacol. 89: 485-491.
18. Pan AC, Borhani DW, Dror RO, Shaw DE, (2013), Molecular determinants of drug-receptor binding kinetics, Drug Discov Today. 18: 667-673.
19. Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, Xu H, Borhani DW, Shaw DE, (2011), Activation mechanism of the beta2-adrenergic receptor, Proc Natl Acad Sci U S A. 108: 18684-18689.

20. Mason JS, Bortolato A, Weiss DR, Deflorian F, Tehan B, Marshall FH, (2013), High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hotspots and explicit water networks, In Silico Pharmacology. 1: 23.
21. Heifetz A, James T, Morao I, Bodkin MJ, Biggin PC, (2016), Guiding lead optimization with GPCR structure modeling and molecular dynamics, Curr Opin Pharmacol. 30: 14-21.
22. Deprez-Poulain R, Deprez B, (2004), Facts, figures and trends in lead generation, Curr Top Med Chem. 4: 569-580.
23. Heifetz A, Aldeghi M, Chudyk E, Fedorov DG, Bodkin M, Biggin PC, (2016), Using the Fragment Molecular Orbital Method to Investigate Agonist-Orexin 2 Receptor Interactions Biochemical Society Transactions
24. Heifetz A, Chudyk EI, Gleave L, Aldeghi M, Cherezov V, Fedorov DG, Biggin PC, Bodkin MJ, (2016), The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR-Ligand Interactions, J Chem Inf Model. 56: 159-172.
25. Heifetz A, Storer RI, McMurray G, James T, Morao I, Aldeghi M, Bodkin MJ, Biggin PC, (2016), Application of an Integrated GPCR SAR-Modeling Platform To Explain the Activation Selectivity of Human 5-HT over 5-HT, ACS Chem Biol.
26. Storer RI, Brennan PE, Brown AD, Bungay PJ, Conlon KM, Corbett MS, DePianta RP, Fish PV, Heifetz A, Ho DK, Jessiman AS, McMurray G, de Oliveira CA, Roberts LR, Root JA, Shanmugasundaram V, Shapiro MJ, Skerten M, Westbrook D, Wheeler S, Whitlock GA, Wright J, (2014), Multiparameter optimization in CNS drug discovery: design of pyrimido[4,5-d]azepines as potent 5-hydroxytryptamine 2C (5-HT(2)C) receptor agonists with exquisite functional selectivity over 5-HT(2)A and 5-HT(2)B receptors, J Med Chem. 57: 5258-5269.
27. Tautermann CS, (2014), GPCR structures in drug design, emerging opportunities with new structures, Bioorg Med Chem Lett. 24: 4073-4079.
28. Bartuzi D, Kaczor AA, Targowska-Duda KM, Matosiuk D, (2017), Recent Advances and Applications of Molecular Docking to G Protein-Coupled Receptors, Molecules. 22.
29. Kitchen DB, Decornez H, Furr JR, Bajorath J, (2004), Docking and scoring in virtual screening for drug discovery: methods and applications, Nat Rev Drug Discov. 3: 935-949.
30. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ, (1998), Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, Journal of Computational Chemistry. 19: 1639-1662.
31. Trott O, Olson AJ, (2010), AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J Comput Chem. 31: 455-461.
32. Rarey M, Kramer B, Lengauer T, Klebe G, (1996), A fast flexible docking method using an incremental construction algorithm, J Mol Biol. 261: 470-489.
33. Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD, (2003), Improved protein-ligand docking using GOLD, Proteins. 52: 609-623.
34. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS, (2004), Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, J Med Chem. 47: 1739-1749.
35. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE, 3rd, (2000), Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, Acc Chem Res. 33: 889-897.
36. Liu S, Wu Y, Lin T, Abel R, Redmann JP, Summa CM, Jaber VR, Lim NM, Mobley DL, (2013), Lead Optimization Mapper: Automating free energy calculations for lead optimization, Journal of computer-aided molecular design. 27: 10.1007/s10822-10013-19678-y.

37. Sotriffer CA, Flader W, Winger RH, Rode BM, Liedl KR, Varga JM, (2000), Automated docking of ligands to antibodies: methods and applications, Methods. 20: 280-291.
38. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ, (2009), AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility, J Comput Chem. 30: 2785-2791.
39. Blundell CD, Packer MJ, Almond A, (2013), Quantification of free ligand conformational preferences by NMR and their relationship to the bioactive conformation, Bioorg Med Chem. 21: 4976-4987.
40. Hawkins PC, Skillman AG, Nicholls A, (2007), Comparison of shape-matching and docking as virtual screening tools, J Med Chem. 50: 74-82.
41. Marino KA, Shang Y, Filizola M, (2017), Insights into the Function of Opioid Receptors from Molecular Dynamics Simulations of Available Crystal Structures, Br J Pharmacol.
42. Schneider S, Provasi D, Filizola M, (2015), The Dynamic Process of Drug-GPCR Binding at Either Orthosteric or Allosteric Sites Evaluated by Metadynamics, Methods Mol Biol. 1335: 277-294.
43. Kaczor AA, Rutkowska E, Bartuzi D, Targowska-Duda KM, Matosiuk D, Selent J, (2016), Computational methods for studying G protein-coupled receptors (GPCRs), Methods Cell Biol. 132: 359-399.
44. Bartuzi D, Kaczor AA, Matosiuk D, (2015), Activation and Allosteric Modulation of Human mu Opioid Receptor in Molecular Dynamics, J Chem Inf Model. 55: 2421-2434.
45. Labute P, (2010), LowModeMD--implicit low-mode velocity filtering applied to conformational search of macrocycles and protein loops, J Chem Inf Model. 50: 792-800.
46. De Vivo M, Masetti M, Bottegoni G, Cavalli A, (2016), Role of Molecular Dynamics and Related Methods in Drug Discovery, J Med Chem. 59: 4035-4061.
47. Mollica L, Theret I, Antoine M, Perron-Sierra F, Charton Y, Fourquez J-M, Wierzbicki M, Boutin JA, Ferry G, Decherchi S, Bottegoni G, Ducrot P, Cavalli A, (2016), Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein–Ligand Residence Times, Journal of Medicinal Chemistry. 59: 7167-7176.
48. Copeland RA, (2016), The drug-target residence time model: a 10-year retrospective, Nat Rev Drug Discov. 15: 87-95.
49. Heifetz A, Trani G, Aldeghi M, MacKinnon CH, McEwan PA, Brookfield FA, Chudyk E, Bodkin M, Pei Z, Burch JD, Ortwine DF, (2016), Fragment Molecular Orbital Method Applied to Lead Optimization of Novel Interleukin-2 Inducible T-Cell Kinase (ITK) Inhibitors J Med Chem.
50. Nishimoto Y, Fedorov DG, (2017), Three-body expansion of the fragment molecular orbital method combined with density-functional tight-binding, J Comput Chem. 38: 406-418.
51. Heifetz A, Aldeghi M, Chudyk EI, Fedorov DG, Bodkin MJ, Biggin PC, (2016), Using the fragment molecular orbital method to investigate agonist-orexin-2 receptor interactions, Biochem Soc Trans. 44: 574-581.
52. Abel R, Young T, Farid R, Berne BJ, Friesner RA, (2008), Role of the active-site solvent in the thermodynamics of factor Xa ligand binding, J Am Chem Soc. 130: 2817-2831.
53. Ross GA, Morris GM, Biggin PC, (2012), Rapid and accurate prediction and scoring of water molecules in protein binding sites, PLoS One. 7: e32036.
54. Truchon JF, Pettitt BM, Labute P, (2014), A Cavity Corrected 3D-RISM Functional for Accurate Solvation Free Energies, J Chem Theory Comput. 10: 934-941.
55. Gerogiokas G, Southey MW, Mazanetz MP, Heifetz A, Bodkin M, Law RJ, Henchman RH, Michel J, (2016), Assessment of Hydration Thermodynamics at Protein Interfaces with Grid Cell Theory, J Phys Chem B. 120: 10442-10452.

56. Gerogiokas G, Southey MW, Mazanetz MP, Heifetz A, Bodkin M, Law RJ, Michel J, (2015), Evaluation of water displacement energetics in protein binding sites with grid cell theory, Phys Chem Chem Phys. 17: 8416-8426.
57. Vajda S, Guarnieri F, (2006), Characterization of protein-ligand interaction sites using experimental and computational methods, Curr Opin Drug Discov Devel. 9: 354-362.
58. Goldmann D, Zdrazil B, Digles D, Ecker GF, (2016), Empowering pharmacoinformatics by linked life science data, J Comput Aided Mol Des.
59. Mazanetz MP, Marmon RJ, Reisser CB, Morao I, (2012), Drug discovery applications for KNIME: an open source data mining platform, Curr Top Med Chem. 12: 1965-1979.
60. Heifetz A, Barker O, Verquin G, Wimmer N, Meutermans W, Pal S, Law RJ, Whittaker M, (2013), Fighting obesity with a sugar-based library: discovery of novel MCH-1R antagonists by a new computational-VAST approach for exploration of GPCR binding sites, J Chem Inf Model. 53: 1084-1099.
61. Tye H, Mueller SG, Prestle J, Scheuerer S, Schindler M, Nosse B, Prevost N, Brown CJ, Heifetz A, Moeller C, Pedret-Dunn A, Whittaker M, (2011), Novel 6,7,8,9-tetrahydro-5H-1,4,7,10a-tetraaza-cyclohepta[f]indene analogues as potent and selective 5-HT(2C) agonists for the treatment of metabolic disorders, Bioorg Med Chem Lett. 21: 34-37.
62. Davenport AJ, Moller C, Heifetz A, Mazanetz MP, Law RJ, Ebneth A, Gemkow MJ, (2010), Using electrophysiology and in silico three-dimensional modeling to reduce human Ether-a-go-go related gene K(+) channel inhibition in a histamine H3 receptor antagonist program, Assay Drug Dev Technol. 8: 781-789.
63. Heifetz A, Morris GB, Biggin PC, Barker O, Fryatt T, Bentley J, Hallett D, Manikowski D, Pal S, Reifegerste R, Slack M, Law R, (2012), Study of human Orexin-1 and -2 G-protein-coupled receptors with novel and published antagonists by modeling, molecular dynamics simulations, and site-directed mutagenesis, Biochemistry. 51: 3178-3197.
64. Barnoud J, Monticelli L, (2015), Coarse-grained force fields for molecular simulations, Methods Mol Biol. 1215: 125-149.
65. Gutierrez-de-Teran H, Keranen H, Azuaje J, Rodriguez D, Aqvist J, Sotelo E, (2015), Computer-aided design of GPCR ligands, Methods Mol Biol. 1272: 271-291.
66. Tummino PJ, Copeland RA, (2008), Residence time of receptor-ligand complexes and its effect on biological function, Biochemistry. 47: 5481-5492.
67. Guo D, Hillger JM, AP IJ, Heitman LH, (2014), Drug-target residence time--a case for G protein-coupled receptors, Med Res Rev. 34: 856-892.

Figure legends

Figure 1A. Optimisation cycle for H2L

Figure 1B. Optimisation cycle for LO

Figure 2: A summary schematic of the Hierarchical GPCR Modeling Protocol (HGMP)

Figure 3. HGMP-C4XD workflow

Figure 4. Schematic summary of the FMO approach: (**a**) Workflow for PIEDA calculations and details on each of the PIE terms that are computed (**b**) FMO analysis of human adenosine OX_2 receptor in complex with Suvorexant (PDB ID 4S0V [47]). The carbon atoms of the ligand are shown in light orange and for the receptor are grey. Nitrogen atoms are shown in blue, oxygen in red and chlorine in light green. The fragmented bonds are marked as red discs. The left-hand bar plots describe the sorted PIE of the most significant residues, and the right-hand plots describe the pair interaction energy decomposition analysis (PIEDA) of these key interactions. PIE terms: electrostatics, dispersion, charge-transfer, and exchange-repulsion are color-coded in yellow, blue, red, and green, respectively. The figure is adapted from our previous publication

Figure 5: Example of KNIME workflow

Figure 6: Summary schematic of the VAST-GPCR modeling workflow that led to the discovery of new MCH-1R antagonists

Figure 7. Schematic summarizing how interaction maps derived from GPCR model for potent & selective OX_2 receptor antagonist