Homology Modeling of Transporter proteins

Ingebrigt Sylte, Mari Gabrielsen and Kurt Kristiansen

Molecular Pharmacology and Toxicology, Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway

Corresponding author:

Ingebrigt Sylte, email: Ingebrigt.Sylte@uit.no, Phone: +47 77644705

Department of Medical Biology, Faculty of Health Sciences, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway

Running head: Comparative modelling of carriers and channels

Abstract

Membrane transporter proteins are divided into channels/pores and carriers and constitute protein families of physiological and pharmacological importance. Several presently used therapeutic compounds elucidate their effects by targeting membrane transporter proteins, including anti-arrhythmic, anesthetic, antidepressant, anxiolytic and diuretic drugs. The lack of three-dimensional structures of human transporters hampers experimental studies and drug discovery. In the present chapter, the use of homology modelling for generating structural models of membrane transporter proteins is reviewed. The increasing number of atomic resolution structures available as templates, together with improvements in methods and algorithms for sequence alignments, secondary structure predictions and model generation, in addition to the increase in computational power have increased the applicability of homology modelling for generating structural models of transporter proteins. Different pitfalls and hints for template selection, multiple sequence alignments, generation and optimization, validation of the models, and the use of transporter homology models for structure based virtual ligand screening are discussed.

Key words: Homology modelling, Transporter proteins, Carriers, Channels and pores, Model building and refinements, Model validation, Structure based virtual screening.

1. Introduction

Membrane transporter proteins (channels and carriers) are responsible for cellular extrusion and uptake of ions, electrons, nutrients, signaling molecules, drugs, toxic substances, metabolic products, macromolecules and other components involved in cellular regulation and function. Transporter proteins are necessary for establishing and controlling the voltage gradient across cell membranes and are major determinants for the pharmacokinetics, safety and efficacy of drugs and toxic substances. The Saier group at the University of California, San Diego, has designed and is maintaining the Transporter Classification Database (TCDB; http://www.tcdb.org), which has become an official classification system approved by the International Union of Biochemistry and Molecular Biology [1,2]. The TCDB includes transporter proteins from all types of living organisms, and is organized as a five level hierarchical system of class, subclass, family, subfamily and the particular transporter protein. At present (17th of March, 2020), the TCDB contains 19 634 protein sequences classified into 1449 transporter families based on phylogeny and function. The TCDB also contains PDB-database (https://www.rcsb.org/) codes of known transporter protein structures and links to structural data. Transporters are multi-spanning integral membrane proteins, and most of them form α-helical bundles and/or barrel-like β-hairpin secondary structures [3,4].

Based on the transporter classification system, membrane transporter proteins are divided into channels and carriers. Channels are water channels [2] or ion channels [5] that passively transport substances down an electrochemical gradient (also called facilitated diffusion) with a rapid transportation rate (milliseconds) since multiple molecules can pass the channel simultaneously. The two main types of ion channels are voltage-gated and ligand-gated ion channels. Voltage-gated ion channels are classified according to the ion being translocated, and ions are transported through the channels by diffusion down their electrochemical gradient. Voltage-gated ion channels are commonly targets for anesthetic and anti-arrhythmic drugs.

Ligand-gated ion channels open upon binding of a specific substrate, like the γ-aminobutyric acid (GABA) receptor A ($GABA_AR$), which triggers the opening of a chloride ion selective pore upon GABA binding [6]. The receptor has a pentameric structure of five homologues subunits (a combination of α-, β-, and γ-subunits) surrounding the chloride ion selective pore [7,8]. The $GABA_AR$ is the main target for the benzodiazepines, which function as allosteric modulators of the receptor, and are sedative and anxiolytic drugs. From a pharmacological point of view, ligand-gated ion channels are often classified as ionotropic receptors and not transporters [9].

Carriers show stereospecific substrate specificity, where the binding of the substrate triggers conformational changes that allow movement of the bound substrate and release on the other side of the membrane. They mediate passive transport or active transport against a concentration gradient and comprise, among others, solute carriers [10] and ATP-driven pumps, including ABC transporters [11]. During passive carrier transport, the substrate/solute diffuses along the concentration gradient without consuming energy, while active transport requires energy as the movement of the substrate is against the concentration gradient. Carriers for active transport are divided into primary and secondary active transporters. In primary active transport, hydrolysis of molecules such as ATP provides energy required for transport of a substrate against its concentration gradient. In secondary active transport, the electrochemical gradient generated by migration of ions down the gradient is used to transport substrates against their concentration gradient. The secondary active transport can be by antiporters, where the substrate and ion transport across the membrane is in opposite direction of each other, or by symporters, where the ion and substrate transport is in the same direction. Due to the complicated process of necessary conformational changes, carriers have much lower transportation rates than channels (10^2-10^4 molecules per second for carriers and 10^6-10^7 molecules per second for ion channels) [12].

Examples of carriers include the monoamine transporters of the neurotransmitter: sodium symporter (NSS) family, that belong to the superfamily of solute carriers [13]. The monoamine transporters are secondary transporters expressed in both the central (CNS) and peripheral nervous systems, being responsible for the reuptake of monoamines (5-hydroxytryptamine (5-HT), dopamine and norepinephrine) from the extracellular space into the presynaptic cell. Dysregulation of monoamine mediated synaptic transmission in the CNS is connected to prevalent mental disorders including major depressive disorder (MDD) [14], schizophrenia [15], Parkinson's disease [16] and attention deficit hyperactive disorder (ADHD) [17]. Inhibitors of monoamine transporters are therefore therapeutic agents in the pharmacological treatment of mental disorders, such as the selective serotonin reuptake inhibitors (SSRIs). In addition, monoamine transporters are also the primary sites of action of several psychostimulants and drugs of abuse including cocaine, ecstasy and methamphetamine [13].

Another therapeutically important superfamily of carriers is the ATP-binding cassette (ABC) superfamily, that utilize the energy from hydrolysis of ATP to pump different substrates, including drugs, out of the cells [11]. This superfamily includes the permeability glycoprotein (P-glycoprotein), the multidrug resistance associated protein (MRP1) and the breast cancer resistance protein (BCRP) and several others. Increased expression of these transporters contributes to multidrug resistance (MDR) of multiple structurally unrelated chemotherapeutic drugs [11]. These transporters were first discovered as mediators of MDR, but in addition, they are important for the normal excretion of drugs from the body and in the function of barriers such as the blood-brain barrier (BBB), and are therefore very important for the pharmacokinetics and bioavailability of drugs [11].

In 2011, 67 transport proteins were primary effect-mediating targets for drugs approved by the US Food and Drug Administration (FDA) [18]. This corresponded to 15 % of totally 435 primary effect-mediating targets of FDA approved drugs in 2011 , making transporter proteins

the third most common class of human drug targets after receptors and enzymes. The most common type of transporter drug targets was voltage-gated ion channels with 29 primary effect-mediating targets.

At present (April, 2021), approximately 177.000 structural entities have been deposited in the PDB-database (https://www.rcsb.org/), which is a huge increase from approximately 13.500 entities in 2000 [19]. A recent paper by Goodsell and co-workers show that in July 2019 the PDB-database contained 9834 structures of transporter proteins, of which 4131 were of channels (756 voltage gated-, and 968 ligand-gated ion channels) and pores. The 9834 structures also included accessory factors involved in transport (1651 structures) and incompletely characterized transport systems (952 structures) [19]. The number of available structures of clinical important transporter proteins has also increased, but still quite few human transporters are structurally characterized. The first x-ray crystal structure of an NSS transporter was published in 2005, which was the structure of the sodium dependent leucine transporter LeuT from the bacterium *Aquifex aeolicus* [20], while an inhibitor bound LeuT structure was published in 2008 [21]. The inhibitor bound structure was used, as a template for constructing homology models of human monoamine transporters for several years. The first human NSS transporter structure, the serotonin transporter (SERT), came several years later in 2016 [22]. The increase in the number of deposited structures indicates that technical advances in crystallization and structural data collection by synchrotron and major advances in three dimensional cryo-electron microscopy [23] during the last 10 years have contributed to an increased insight into three-dimensional structures of transporter proteins and other membrane proteins. Especially carriers, but also other transporter proteins are structurally flexible. Capturing the protein structure in interesting conformational states for further studies may be challenging. In spite of the increase in the number of structures, several transporter families are still poorly characterized at the molecular level, despite of clinical significance and potential as

drug targets [24]. The lack of structures at a sufficient level of structural details hampers rational drug design, and limits the understanding of transporter function and interactions.

In lack of 3D structure, homology modeling is a valuable approach for obtaining structural information about transporter proteins. The increase in the number of available templates for homology modelling and improvements in computational power and molecular modelling methods have given increased applicability of homology modelling for generating structural models of transporter proteins. Computational methods for predicting the 3D structures of proteins have been used for several years, and the prediction methods are generally classified as *de novo* modelling, where the 3D structure is predicted directly from its amino acid sequence [25,26], traditional homology modelling (comparative modelling), and treading which mainly is used in combination with one of the other methods [27]. In spite of improved *de novo* methods, the homology modelling approach is still considered as the most accurate approach. A general homology modelling approach can be divided into four steps: 1) Identification of homologous proteins of known structure and selection of the best template or set of templates for the modelling. 2) Generating and optimizing (multiple) sequence alignments between the query sequence and homologues sequences (including template protein sequences). 3) Building and optimizing homology models of the query sequence. 4) Validation of the model(s).

2. Materials

The quality of the amino acid alignment between the template and the target is very important for the quality of the homology model. Methods for generating multiple sequence alignments were originally developed for soluble proteins [28] since knowledge-based reference alignments could be generated based on available 3D structures. Most soluble proteins are globular with a hydrophilic surface while most membrane proteins have quite hydrophobic membrane traversing regions, giving differences in the amino acid substitution preferences

between soluble proteins and membrane proteins. Due to the increase in the number of available structures from different membrane protein families, it has been possible to develop alignments methods for membrane proteins that take into account hydrophobicity profiles and transmembrane region predictions. Examples are the AlignMe program [29] that was developed for sequence alignments of solute carriers structurally resembling the bacterial leucine transporter LeuT, and the MP-T (Membrane Protein Threader) program [30]. Such programs have improved the sequence alignments of membrane proteins and thereby the quality of homology models [30].

A table showing frequently used online servers and software tools for protein homology modelling is given by Muhammad and Aki-Yalcin [31]. These resources are commonly used in the construction of homology models of transporter proteins. We have a long lasting experience with the ICM-modelling program package [32,33], Prime (Schrödinger) [34] and Modeller [35] for generating homology models of solute carriers and G-protein coupled receptors [36-45]. Online servers and services offering automatic generation of homology models are available, like SWISS-MODEL [46], and Phyre^2 [47]. Some online tools are specifically designed for automatic generation of homology models of membrane proteins. These tools have implemented algorithms and methods specifically designed for sequence alignments, prediction of secondary structure, transmembrane regions and 3D models of membrane proteins [48,49]. Examples are MEMOIR [50], MEDELLER [51] and RosettaMembrane [52].

Services for automatically generation of homology models of membrane proteins may produce high quality models, especially when the sequence similarity between template and target is high [53]. Fully automatically generated models may also be decent starting models at low similarity between target and template, and the RosettaMembrane program has been specifically developed for modeling transmembrane helical proteins based on distant homologous as template [52]. At lower similarity, manual adjustments of the different steps in

the modelling procedure are often necessary, like adjustments of the multiple amino acid sequence alignments and introduction of constrains/restrains in building and optimization of the model. Manual adjustments of the alignment can be based on structural superposition of proteins of known 3D structure. Several molecular modelling packages, like the ICM modelling and Schrödinger program packages also contain possibilities for structural superposition of proteins. Most available templates are bacterial, and when using bacterial templates for constructing models of human transporters, each step in the modelling must also be performed with caution due to differences between prokaryotes and eukaryotes. For example, post-translation modifications are lacking in prokaryotes, which may affect protein structure, folding and dynamics and give differences between prokaryotes and eukaryotes. In such cases, a less automatic process where each step is carefully performed, with necessary manual adjustments, may give more accurate models of human transporters than fully automatically generated models [54].

Relevant experimental data to guide the modelling can increase the accuracy of the models, and increase the hit rate during docking and virtual screening. Appropriate experimental data are results from site- directed mutagenesis studies or other molecular biology approaches that can contribute with structural information about the geometry of binding sites or other functionally important protein regions. Information from different biophysical and structural biology studies of transporter structure, function and dynamics can also be important, Further, ligand binding data like substrate specificity and inhibition kinetics, structure activity relationships studies of inhibitors and substrates can be used to refine models and obtain ligand specific (ligand steered) models.

Several programs are available for structural validation of the generated homology models. These program include WhatCheck [55] and PROCHECK [56] that both are using geometrical, stereochemical, and statistical criteria to check the models, and ERRAT [57] which is

comparing the statistics of non-bonded interactions between different atom types of the model and highly refined structures. ICM Protein Health, which is a part of the ICM package, is using normalized force field residue energies and compare the energies with expected energies from high quality crystal structures [58]. In addition, Ramachandran plots can be used to check the backbone geometry of the model. Models can be uploaded to structural validation severs, as the SAVES (https://servicesn.mbi.ucla.edu/SAVES/), and the user can select between different programs for quality checking.

3. Methods

Selection of methods and the reliability of the models relay on the availability of templates close in amino acid sequence and function to the target. Homology models are computationally derived approximations of a protein structure, and will always contain inaccuracies and sometimes errors. The quality required for a model depends largely on its intended use. Low-accuracy models can be completely sufficient for designing mutagenesis experiments, while an overall sequence similarity of more than 50 % between the target and template of soluble proteins is generally believed to be necessary for obtaining models that can be used for structure based drug discovery [59]. For mechanistic studies, the highest possible level of accuracy is essential [60]. The transmembrane regions and ligand binding sites are highly conserved within membrane protein families, despite the fact that the overall sequence similarity can be much lower than 50 %. Forrest and co-workers showed that a sequence similarity of approximately 30 % in transmembrane regions between template and target gave a Cα root mean square deviation (RMSD) of 2 Å in these regions between the model and x-ray structure template [61].

A flow chart indicating the main steps in a homology modelling procedure of transporter proteins is shown in Fig. 1. The following sections outline particular steps in the scheme.

3.1 Template identification and selection

Template selection is most often based on a traditional BLAST (Basic Local Alignment Search Tool) search for identifying templates most similar in sequence to the query sequence (the template). The structure closest in sequence to the target sequence is then most often used as a template for modelling the target. However, using the sequence similarity only as the criterion for template selection will not always give the most optimal model (see **Note 1**).

Carriers undergo substantial conformational changes during the transport cycle. For solute carriers, an "alternating access" mechanism has been proposed that requires transition between at least three conformational states in which the ion and substrate binding sites are alternately exposed to the inner and outer side of the membrane, or occluded within the carrier [62]. Crystal structures of the sodium dependent leucine transporter LeuT from the bacterium *Aquifex aeolicus* [63], and other secondary transporters with similar 3D fold [64,65], support the suggested translocation mechanisms, and the structures are classified as being in inward-facing, outward-facing or substrate-occluded states describing the putative pathway for substrate binding, translocation and release. Additional elucidation of the "alternating access" mechanism was given by biophysical studies and theoretical calculations [66,67]. Several energetically stable conformational states may therefore be possible during the transport cycle, which also needs to be taken into account during template selection. The putative template closest in sequence to the target sequence is not necessarily the best template, since it may not represent the conformational state in the transport cycle that was intended to model.

Transporter proteins are structurally flexible, and during binding, the ligand binding cavity adopts to the structure of the binding molecule. The binding site structure may therefore be very different between an *apo* and *holo* structures of a transporter protein (**Note 2**). Due to structural adaption, the *holo* structure may be quite selective for the particular complexed ligand or for a

structural group of ligands. Binding site differences between *apo* and *holo* structures also need to be taken into account during template selection and in docking experiments.

Models of transporter proteins are often constructed based on templates with relative low levels of overall sequence identity with the target (less than 30 %), due to the lack of available templates with higher similarity [54]. The application of such models may be limited. However, functionally important regions like substrate binding sites exhibit most often higher degrees of conservation than the rest of the structure, and fairly accurate structural models of binding sites may still be constructed in spite of poor overall accuracy [59]. It has been shown, by us and others, that homology models with an overall similarity of less than 25 % between the template and target may successfully be used for structure based virtual ligand screening. For example by using homology models of the noradrenaline transporter (NET) and experimental verification, Schlessinger and coworkers identified NET inhibitors [68], while we used homology models of the SERT and experimental verification to identify SERT inhibitors [37]. In both studies, the structure based virtual ligand screening was performed with homology models based on the structure of the leucine transporter LeuT from the bacterium *Aquifex aeolicus* [20,21]. The overall amino acid similarity between LeuT and SERT is 21 %, while the similarity in transmembrane helices involved in substrate binding is 35 % [38]. Fig 2 shows the binding site of LeuT with the inhibitor L-Trp [21], which was used as template for our homology models, and the binding site of human SERT with the SSRI paroxetine [22], which was determined after our homology models. These x-ray structures show that there are both conserved and non-conserved amino acids between the binding sites, and that the L-Trp binding site in LeuT is narrower than the paroxetine binding site in SERT, which may indicate that LeuT has adopted to the smaller inhibitor. Our initial LeuT-based SERT models could not dock most of the known SERT inhibitors [38], and special treatment of the binding site was necessary. By using the ICM software to generate multiple conformations of the binding site

and perform ensemble docking [38,69,70], we were able to dock most known SERT inhibitors and select binding site conformations for structure based virtual ligand screening that recognized new SERT inhibitors [37]. These studies indicate that special treatment of binding site amino acids may be important for the success rate of docking and structure based virtual ligand screening when using carrier models based on low sequence identify between template and target (**Note 3**).

A single model based on one template only represents a static snapshot, which will reduce the feasibility of the generated model. If several templates are available for the transporter, several models can be constructed, and in that way structural flexibility is partly taken into account. Another relevant factor for template selection is the structural quality of template structures. In general, high resolution structural templates should be favored over low resolution templates. Several putative template structures for transporter proteins are low resolution structures from cryo-electron microscopy. Experimental conditions, putative structural errors and crystal packing forces will also affect the quality of homology models, and should also be considered in the selection of templates. If crystal packing forces affect particular interesting areas it will affect the quality of the homology model.

3.2 Target-template alignments

The sequence alignment between target and template is as a very critical step, and the quality of the alignment determines the quality of the model. Small mistakes in the alignment may give limited accuracy of the models. A multiple sequence alignment is recommended as a basis for the alignment between target and templates. Such an alignment will highlight evolutionary relationships within the family and increase the probability that corresponding sequence positions are correctly aligned. However, the sequences used in the alignment should be carefully inspected such that the sequence conservation is not biased towards a subset of

sequences within the family/subfamily (**Note 4**). If more than one temple is available, it may also be useful to adjust the alignment based on structural superposition of the known structures (**Note 5**).

The amino acid similarity between template and target may be low which complicates the alignment procedure. For soluble proteins, an overall sequence identity between template and query higher than 40% will normally give an alignment with few gaps, and models where approximately 90 % of backbone atoms can be modeled with an RMSD of about 1 Å. An overall sequence identity of 30 % - 40 % will normally results in more frequent insertions and deletion in the alignment and models where 80 % of backbone atoms are modelled with an RMSD of approximately 3.5 Å [71,72]. For membrane protein families the overall amino acids similarity may be quite low, but the structure of membrane spanning areas and binding sites for endogenous activators are well conserved.

3.3 Model building and refinements

The model building of transporter homology models briefly involves three main steps: 1. Construction the structurally conserved core region, which for most transporters are the transmembrane parts. 2. Construction of extracellular and intracellular loop regions, which normally are the less conserved parts. 3. Optimization of side chains conformation and energy refinements of the model.

The methods used for construction of conserved core regions can be classified into rigid body-assembly methods [33], segment matching methods [73], spatial restraint methods [35] and artificial evolution methods [74,75]. Reviews of core construction methods used by the most popular homology modelling programs are given by Xiang [71] and Muhammad and Aki-Yalcin [31]. The ICM program [33] and Prime (Schrödinger) [34] use the rigid body assembly

method, while MODELLER which is the most popular homology modelling program (based on citations) uses the spatial restraint method [35]. Some programs use a combination of different methods for constructing the core regions [76].

The length of loops and terminals may differ substantially within a membrane protein family, and the construction of these regions are therefore much more uncertain than of the conserved core regions. The structure of terminals and loops may be important for binding specificity and function, and the accuracy of these parts is an important factor for the application of the model in further studies. The inclusion of loops in the model may therefore depend on the planned application of the model, and wrongly modelled loop structure may induce structural stress into conserved regions during refinements (**Note 6**). For non-conserved shorter loops (4-7 amino acids), a database loop search is most often used, where available structures in the PDB are searched to provide the loop structure. Another approach is to use *de novo* prediction methods to search the conformational space of the loop. Monte Carlo (MC) simulations, molecular dynamics (MD) simulations, simulated annealing and genetic algorithms are used, and often, in combination [77-79].

The model refinements process usually involves removal of clashes and geometrical regularisation of bond lengths and angles, but may also involve more sophisticated structural corrections. The refinements process may be performed with traditional molecular mechanical force field programs and often starts with an energy minimization, and involve different steps of side chain conformational sampling, interactive annealing of backbone atoms, and refinements by MC and MD simulations. The different steps in the refinements may eliminate structural errors, but it is important to have in mind that other error may as well be introduced (**Note 6**).

Structural templates often represent an unliganed state of the binding site (*apo* structure) or the binding site geometry is biased against a particular compound complexed with the template

(*holo* structure). Models based on such templates may also represent unliganed or biased conformations of the binding site. Docking known target binders into these models may be problematic and give low docking score. Incorporating ligand binding data into the model optimization process may improve docking into such models. If the template contains a compound in the binding site, one approach is to replace the compound with a preferred target compound, and treat the ligand as a part of the model through the modelling process, and thereby obtain a ligand gated or ligand steered model with a binding site adjusted to the target compound [80]. A simpler approach is to perform induced fit docking of high affinity compounds that will generate additional conformations of the binding site. Structural clustering of known target compounds and induced fit docking of cluster representatives may give binding site conformations that are specific for a structural cluster of known compounds [44]. Homology models optimized by docking of known compounds (ligand-based models) may give an improved accuracy of the binding site conformation, and have been shown to increase docking enrichment [81], and increase the hit rate during structure based virtual screening experiments [37,59]. However, the success demands on correct docking of compounds used to optimize the homology models.

3.4 Model validation

Homology models of channels and carriers will always contain uncertainties and shortcomings, especially when the similarity in sequence and function between the template and target is low. However, models generated from templates of low similarity may still be used as a working tool for generating hypothesis and designing site directed mutagenesis.

The models need to be validated both for spatial feasibility and predictive applicability. Evaluation of spatial feasibility may access local and global structural errors and may be based on geometrical, stereochemical, statistical and/or energy criteria. The validation can form the

basis for additional refinements of the model, adjustments of the target-template alignment and rebuilding models. Predicting transporter-ligand interactions by using structure based virtual ligand screening enrichment has become a commonly used approach for testing the model compliance with experimental ligand binding data. A dataset consisting of known potent compounds for the target and decoys (typical ration of 1:50) is docked and the compounds are ranked by predicted binding affinities. The decoys resemble potent binders in molecular weight, number of atoms, and physiochemical properties, and are presumed non-binders. If the model is capable of scoring the known binders in front of the decoys, the model is considered to be predictive and has a good potential for structure based drug discovery. This approach can be used to evaluate between models before a virtual screening campaign (**Note 7**).

The optimal testing of a model or a set of model is to design and perform *in vitro* studies based on the models. Several iterative cycles of spatial feasibility and predictivity testing and models adjustments may be necessary before *in vitro* testing. The *in vitro* testing may for example be site directed mutagenesis combined by ligand binding studies, or testing of hits from a structure based virtual screening campaign.

4 Notes

1. A careful selection between appropriate templates is necessary. The template closest in sequence to the target is not necessary the most appropriate. Most transporter proteins (especially carriers) undergo substantial conformational changes during the transport cycle and the template conformation may not be in an appropriate conformation for the target model.
2. If the purpose is to study transporter-ligand interactions, it must be considered if templates represent conformations biased for a ligand or a group of ligands (*apo* or *holo* structures).

3. Proper treatment of the binding site amino acids are important for the success rate when docking into carrier models based on low sequence identify between template and target. A docking protocol taking the structural flexibility of the binding site (induced fit, ensemble docking) may increase the success rate.
4. The sequences used in the multiple sequence alignment should be carefully selected to avoid that the sequence conservation is biased towards a subset of sequences within the family/subfamily. However, sequences lacking known 3D structures should also be included in the multiple sequence alignment, since that will highlight the family/subfamily sequence conservation.
5. Manual adjustments of the multiple sequence alignment may be necessary. Such and alignment must be based on a structural superimposition of known structures as a knowledge-based reference for the alignment.
6. Homology models of transporters need careful refinements. A tough global refinement process using molecular mechanics force field programs for energetically and structurally refinements by MD or MC simulations may as well induce uncertainties into conserved regions. This is particularly important if the homology between the template and target is low, and binding sites for ions and water molecules are not conserved. MD and MC simulations for refinements may than induce structural stress into conserved regions.
7. The ligand datasets used to validate models by docking should not be biased against a subset of the known binders for the target.

5 References

1. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G (2016) The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res 44 (D1):D372-379. doi:10.1093/nar/gkv1103

2. Brown D (2017) The Discovery of Water Channels (Aquaporins). Ann Nutr Metab 70 Suppl 1:37-42. doi:10.1159/000463061

3. Vinothkumar KR, Henderson R (2010) Structures of membrane proteins. Q Rev Biophys 43 (1):65-158. doi:10.1017/S0033583510000041

4. Niitsu A, Heal JW, Fauland K, Thomson AR, Woolfson DN (2017) Membrane-spanning alpha-helical barrels as tractable protein-design targets. Philos Trans R Soc Lond B Biol Sci 372 (1726). doi:10.1098/rstb.2016.0213

5. Liu Y, Wang K (2019) Exploiting the Diversity of Ion Channels: Modulation of Ion Channels for Therapeutic Indications. Handb Exp Pharmacol 260:187-205. doi:10.1007/164_2019_333

6. Sigel E, Steinmann ME (2012) Structure, function, and modulation of GABA(A) receptors. J Biol Chem 287 (48):40224-40231. doi:10.1074/jbc.R112.386664

7. Masiulis S, Desai R, Uchanski T, Martin IS, Laverty D, Karia D, Malinauskas T, Zivanov J, Pardon E, Kotecha A, Steyaert J, Miller KW, Aricescu AR (2019) GABA(A) receptor signalling mechanisms revealed by structural pharmacology. Nature 565 (7740):454-459.

8. Laverty D, Desai R, Uchanski T, Masiulis S, Stec WJ, Malinauskas T, Zivanov J, Pardon E, Steyaert J, Miller KW, Aricescu AR (2019) Cryo-EM structure of the human alpha 1 beta 3 gamma 2 GABA(A) receptor in a lipid bilayer. Nature 565 (7740):516-520.

9. Alexander SP, Peters JA, Kelly E, Marrion NV, Faccenda E, Harding SD, Pawson AJ, Sharman JL, Southan C, Davies JA, Collaborators C (2017) THE CONCISE GUIDE TO

PHARMACOLOGY 2017/18: Ligand-gated ion channels. Br J Pharmacol 174 Suppl 1:S130-S159. doi:10.1111/bph.13879

10. Cesar-Razquin A, Snijder B, Frappier-Brinton T, Isserlin R, Gyimesi G, Bai X, Reithmeier RA, Hepworth D, Hediger MA, Edwards AM, Superti-Furga G (2015) A Call for Systematic Research on Solute Carriers. Cell 162 (3):478-487. doi:10.1016/j.cell.2015.07.022

11. Lusvarghi S, Robey RW, Gottesman MM, Ambudkar SV (2020) Multidrug transporters: recent insights from cryo-electron microscopy-derived atomic structures and animal models. F1000Res 9. doi:10.12688/f1000research.21295.1

12. Friedmann MH (2008) Facilitated diffusion: channels and carriers. In: Principles and models of biological transport. Springer, pp 111-179

13. Aggarwal S, Mortensen OV (2017) In Vitro Assays for the Functional Characterization of the Dopamine Transporter (DAT). Curr Protoc Pharmacol 79:12.17.11-12.17.21. doi:10.1002/cpph.33

14. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, Mohr DC, Schatzberg AF (2016) Major depressive disorder. Nat Rev Dis Primers 2:16065. doi:10.1038/nrdp.2016.65

15. Takano H (2018) Cognitive Function and Monoamine Neurotransmission in Schizophrenia: Evidence From Positron Emission Tomography Studies. Front Psychiatry 9:228. doi:10.3389/fpsyt.2018.00228

16. German CL, Baladi MG, McFadden LM, Hanson GR, Fleckenstein AE (2015) Regulation of the Dopamine and Vesicular Monoamine Transporters: Pharmacological Targets and Implications for Disease. Pharmacol Rev 67 (4):1005-1024. doi:10.1124/pr.114.010397

17. Faraone SV (2018) The pharmacology of amphetamine and methylphenidate: Relevance to the neurobiology of attention-deficit/hyperactivity disorder and other psychiatric comorbidities. Neurosci Biobehav Rev 87:255-270. doi:10.1016/j.neubiorev.2018.02.001

18. Rask-Andersen M, Almen MS, Schioth HB (2011) Trends in the exploitation of novel drug targets. Nat Rev Drug Discov 10 (8):579-590. doi:10.1038/nrd3478

19. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, Segura J, Shao C, Voigt M, Westbrook JD, Young JY, Burley SK (2020) RCSB Protein Data Bank: Enabling biomedical research and drug discovery. Protein Sci 29 (1):52-65. doi:10.1002/pro.3730

20. Yamashita A, Singh SK, Kawate T, Jin Y, Gouaux E (2005) Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature 437 (7056):215-223. doi:10.1038/nature03978

21. Singh SK, Piscitelli CL, Yamashita A, Gouaux E (2008) A competitive inhibitor traps LeuT in an open-to-out conformation. Science 322 (5908):1655-1661. doi:10.1126/science.1166777

22. Coleman JA, Green EM, Gouaux E (2016) X-ray structures and mechanism of the human serotonin transporter. Nature 532 (7599):334-339. doi:10.1038/nature17629

23. Garcia-Nafria J, Tate CG (2020) Cryo-Electron Microscopy: Moving Beyond X-Ray Crystal Structures for Drug Receptors and Drug Development. Annu Rev Pharmacol Toxicol 60:51-71. doi:10.1146/annurev-pharmtox-010919-023545

24. Hediger MA, Clemencon B, Burrier RE, Bruford EA (2013) The ABCs of membrane transporters in health and disease (SLC series): introduction. Mol Aspects Med 34 (2-3):95-107. doi:10.1016/j.mam.2012.12.009

25. Mullins JG (2012) Structural modelling pipelines in next generation sequencing projects. Adv Protein Chem Struct Biol 89:117-167. doi:10.1016/B978-0-12-394287-6.00005-7

26. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A (2016) Coarse-Grained Protein Models and Their Applications. Chem Rev 116 (14):7898-7936. doi:10.1021/acs.chemrev.6b00163

27. Casadio R, Fariselli P, Martelli PL, Tasco G (2007) Thinking the impossible: how to solve the protein folding problem with and without homologous structures and more. Methods Mol Biol 350:305-320. doi:10.1385/1-59745-189-4:305

28. Venclovas C (2012) Methods for sequence-structure alignment. Methods Mol Biol 857:55-82. doi:10.1007/978-1-61779-588-6_3

29. Stamm M, Staritzbichler R, Khafizov K, Forrest LR (2014) AlignMe--a membrane protein sequence alignment web server. Nucleic Acids Res 42 (Web Server issue):W246-251. doi:10.1093/nar/gku291

30. Hill JR, Deane CM (2013) MP-T: improving membrane protein alignment for structure prediction. Bioinformatics 29 (1):54-61. doi:10.1093/bioinformatics/bts640

31. Muhammed MT, Aki-Yalcin E (2019) Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chem Biol Drug Des 93 (1):12-20. doi:10.1111/cbdd.13388

32. Cardozo T, Totrov M, Abagyan R (1995) Homology modeling by the ICM method. Proteins 23 (3):403-414. doi:10.1002/prot.340230314

33. Abagyan R, Totrov M, Kuznetsov D (1994) Icm - a New Method for Protein Modeling and Design - Applications to Docking and Structure Prediction from the Distorted Native Conformation. J Comput Chem 15 (5):488-506

34. Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, Friesner RA (2004) A hierarchical approach to all-atom protein loop prediction. Proteins-Structure Function and Bioinformatics 55 (2):351-367

35. Sali A, Blundell TL (1993) Comparative Protein Modeling by Satisfaction of Spatial Restraints. J Mol Biol 234 (3):779-815

36. Gabrielsen M, Ravna AW, Kristiansen K, Sylte I (2012) Substrate binding and translocation of the serotonin transporter studied by docking and molecular dynamics simulations. J Mol Model 18 (3):1073-1085. doi:10.1007/s00894-011-1133-1

37. Gabrielsen M, Kurczab R, Siwek A, Wolak M, Ravna AW, Kristiansen K, Kufareva I, Abagyan R, Nowak G, Chilmonczyk Z, Sylte I, Bojarski AJ (2014) Identification of novel serotonin transporter compounds by virtual screening. J Chem Inf Model 54 (3):933-943. doi:10.1021/ci400742s

38. Gabrielsen M, Kurczab R, Ravna AW, Kufareva I, Abagyan R, Chilmonczyk Z, Bojarski AJ, Sylte I (2012) Molecular mechanism of serotonin transporter inhibition elucidated by a new flexible docking protocol. Eur J Med Chem 47 (1):24-37. doi:10.1016/j.ejmech.2011.09.056

39. Warszycki D, Rueda M, Mordalski S, Kristiansen K, Satala G, Rataj K, Chilmonczyk Z, Sylte I, Abagyan R, Bojarski AJ (2017) From Homology Models to a Set of Predictive Binding Pockets-a 5-HT1A Receptor Case Study. J Chem Inf Model 57 (2):311-321. doi:10.1021/acs.jcim.6b00263

40. Ravna AW, Sylte I, Sager G (2009) Binding site of ABC transporter homology models confirmed by ABCB1 crystal structure. Theor Biol Med Model 6:20. doi:10.1186/1742-4682-6-20

41. Ravna AW, Sylte I, Kristiansen K, Dahl SG (2006) Putative drug binding conformations of monoamine transporters. Bioorg Med Chem 14 (3):666-675. doi:10.1016/j.bmc.2005.08.054

42. Jaronczyk M, Wolosewicz K, Gabrielsen M, Nowak G, Kufareva I, Mazurek AP, Ravna AW, Abagyan R, Bojarski AJ, Sylte I, Chilmonczyk Z (2012) Synthesis, in vitro binding studies and docking of long-chain arylpiperazine nitroquipazine analogues, as potential serotonin transporter inhibitors. Eur J Med Chem 49:200-210. doi:10.1016/j.ejmech.2012.01.012

43. Gabrielsen M, Wolosewicz K, Zawadzka A, Kossakowski J, Nowak G, Wolak M, Stachowicz K, Siwek A, Ravna AW, Kufareva I, Kozerski L, Bednarek E, Sitkowski J, Bocian W, Abagyan R, Bojarski AJ, Sylte I, Chilmonczyk Z (2013) Synthesis, antidepressant evaluation and docking studies of long-chain alkylnitroquipazines as serotonin transporter inhibitors. Chem Biol Drug Des 81 (6):695-706. doi:10.1111/cbdd.12116

44. Freyd T, Warszycki D, Mordalski S, Bojarski AJ, Sylte I, Gabrielsen M (2017) Ligand-guided homology modelling of the GABAB2 subunit of the GABAB receptor. PLoS One 12 (3):e0173889. doi:10.1371/journal.pone.0173889

45. Baglo Y, Gabrielsen M, Sylte I, Gederaas OA (2013) Homology modeling of human gamma-butyric acid transporters and the binding of pro-drugs 5-aminolevulinic acid and methyl aminolevulinic acid used in photodynamic therapy. PLoS One 8 (6):e65200. doi:10.1371/journal.pone.0065200

46. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46 (W1):W296-W303. doi:10.1093/nar/gky427

47. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10 (6):845-858. doi:10.1038/nprot.2015.053

48. Venko K, Roy Choudhury A, Novic M (2017) Computational Approaches for Revealing the Structure of Membrane Transporters: Case Study on Bilitranslocase. Comput Struct Biotechnol J 15:232-242. doi:10.1016/j.csbj.2017.01.008

49. Almeida JG, Preto AJ, Koukos PI, Bonvin A, Moreira IS (2017) Membrane proteins structures: A review on computational modeling tools. Biochim Biophys Acta Biomembr 1859 (10):2021-2039. doi:10.1016/j.bbamem.2017.07.008

50. Ebejer JP, Hill JR, Kelm S, Shi J, Deane CM (2013) Memoir: template-based structure prediction for membrane proteins. Nucleic Acids Res 41 (Web Server issue):W379-383. doi:10.1093/nar/gkt331

51. Kelm S, Shi J, Deane CM (2010) MEDELLER: homology-based coordinate generation for membrane proteins. Bioinformatics 26 (22):2833-2840. doi:10.1093/bioinformatics/btq554

52. Chen KY, Sun J, Salvo JS, Baker D, Barth P (2014) High-resolution modeling of transmembrane helical protein structures from distant homologues. PLoS Comput Biol 10 (5):e1003636. doi:10.1371/journal.pcbi.1003636

53. Nikolaev DM, Shtyrov AA, Panov MS, Jamal A, Chakchir OB, Kochemirovsky VA, Olivucci M, Ryazantsev MN (2018) A Comparative Study of Modern Homology Modeling Algorithms for Rhodopsin Structure Prediction. Acs Omega 3 (7):7555-7566

54. Schlessinger A, Welch MA, van Vlijmen H, Korzekwa K, Swaan PW, Matsson P (2018) Molecular Modeling of Drug-Transporter Interactions-An International Transporter Consortium Perspective. Clin Pharmacol Ther 104 (5):818-835. doi:10.1002/cpt.1174

55. Hooft RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. Nature 381 (6580):272. doi:10.1038/381272a0

56. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr 26:283-291

57. Colovos C, Yeates TO (1993) Verification of Protein Structures - Patterns of Nonbonded Atomic Interactions. Protein Sci 2 (9):1511-1519

58. Maiorov V, Abagyan R (1998) Energy strain in three-dimensional protein structures. Fold Des 3 (4):259-269

59. Schmidt T, Bergner A, Schwede T (2014) Modelling three-dimensional protein structures for applications in drug design. Drug Discov Today 19 (7):890-897. doi:10.1016/j.drudis.2013.10.027

60. Baker D, Sali A (2001) Protein structure prediction and structural genomics. Science 294 (5540):93-96. doi:10.1126/science.1065659

61. Forrest LR, Tang CL, Honig B (2006) On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. Biophys J 91 (2):508-517

62. Jardetzky O (1966) Simple allosteric model for membrane pumps. Nature 211 (5052):969-970. doi:10.1038/211969a0

63. Krishnamurthy H, Gouaux E (2012) X-ray structures of LeuT in substrate-free outward-open and apo inward-open states. Nature 481 (7382):469-474. doi:10.1038/nature10737

64. Shimamura T, Weyand S, Beckstein O, Rutherford NG, Hadden JM, Sharples D, Sansom MS, Iwata S, Henderson PJ, Cameron AD (2010) Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. Science 328 (5977):470-473. doi:10.1126/science.1186303

65. Perez C, Koshy C, Yildiz O, Ziegler C (2012) Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. Nature 490 (7418):126-130. doi:10.1038/nature11403

66. Khelashvili G, Stanley N, Sahai MA, Medina J, LeVine MV, Shi L, De Fabritiis G, Weinstein H (2015) Spontaneous inward opening of the dopamine transporter is triggered by PIP2-regulated dynamics of the N-terminus. ACS Chem Neurosci 6 (11):1825-1837. doi:10.1021/acschemneuro.5b00179

67. Kazmier K, Sharma S, Quick M, Islam SM, Roux B, Weinstein H, Javitch JA, McHaourab HS (2014) Conformational dynamics of ligand-dependent alternating access in LeuT. Nat Struct Mol Biol 21 (5):472-479. doi:10.1038/nsmb.2816

68. Schlessinger A, Geier E, Fan H, Irwin JJ, Shoichet BK, Giacomini KM, Sali A (2011) Structure-based discovery of prescription drugs that interact with the norepinephrine transporter, NET. Proc Natl Acad Sci U S A 108 (38):15810-15815. doi:10.1073/pnas.1106030108

69. Rueda M, Bottegoni G, Abagyan R (2009) Consistent improvement of cross-docking results using binding site ensembles generated with elastic network normal modes. J Chem Inf Model 49 (3):716-725. doi:10.1021/ci8003732

70. Bottegoni G, Kufareva I, Totrov M, Abagyan R (2009) Four-dimensional docking: a fast and accurate account of discrete receptor flexibility in ligand docking. J Med Chem 52 (2):397-406. doi:10.1021/jm8009958

71. Xiang Z (2006) Advances in homology protein structure modeling. Curr Protein Pept Sci 7 (3):217-227. doi:10.2174/138920306777452312

72. Sauder JM, Arthur JW, Dunbrack RL, Jr. (2000) Large-scale comparison of protein sequence alignment algorithms with structure alignments. Proteins 40 (1):6-22. doi:10.1002/(sici)1097-0134(20000701)40:1<6::aid-prot30>3.0.co;2-7

73. Levitt M (1992) Accurate Modeling of Protein Conformation by Automatic Segment Matching. J Mol Biol 226 (2):507-533

74. Xiang ZX, Soto CS, Honig B (2002) Evaluating conformational free energies: The colony energy and its application to the problem of loop prediction. P Natl Acad Sci USA 99 (11):7432-7437

75. Xiang ZX, Honig B (2001) Extending the accuracy limits of prediction for side-chain conformations (vol 311, pg 421, 2001). J Mol Biol 312 (2):419-419

76. Dorn M, MB ES, Buriol LS, Lamb LC (2014) Three-dimensional protein structure prediction: Methods and computational strategies. Comput Biol Chem 53PB:251-276. doi:10.1016/j.compbiolchem.2014.10.001

77. Wong SWK, Liu JS, Kou SC (2017) Fast de novo discovery of low-energy protein loop conformations. Proteins 85 (8):1402-1412. doi:10.1002/prot.25300

78. Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K (2009) Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77 Suppl 9:114-122. doi:10.1002/prot.22570

79. Jamroz M, Kolinski A (2010) Modeling of loops in proteins: a multi-method approach. BMC Struct Biol 10:5. doi:10.1186/1472-6807-10-5

80. Dalton JA, Jackson RM (2010) Homology-modelling protein-ligand interactions: allowing for ligand-induced conformational change. J Mol Biol 399 (4):645-661. doi:10.1016/j.jmb.2010.04.047

81. Fan H, Irwin JJ, Webb BM, Klebe G, Shoichet BK, Sali A (2009) Molecular docking screens using comparative models of proteins. J Chem Inf Model 49 (11):2512-2527. doi:10.1021/ci9003706

6 Figure captions

Fig. 1. Flow chart indicating putative steps in the modelling procedure of transporters. Template selection, alignment, model building and refinements (in dark blue), model validation (light blue), and putative use of the final models (green).

Fig. 2. Above: The leucine transporter LeuT from the bacterium *Aquifex aeolicus* in complex with the inhibitor L-Trp (PDB id: 3F3A). Below: The human SERT structure in complex with the inhibitor paroxetine (PDB id:5I6X). Amino acid side chains within 5 Å of the inhibitors have been displayed. Color coding of atoms: oxygen; red, nitrogen; blue, carbon (SERT); light blue, carbon (LeuT); grey, carbon (inhibitors); yellow. Color coding of ions: Na^+: blue sphere, Cl^-: green sphere.

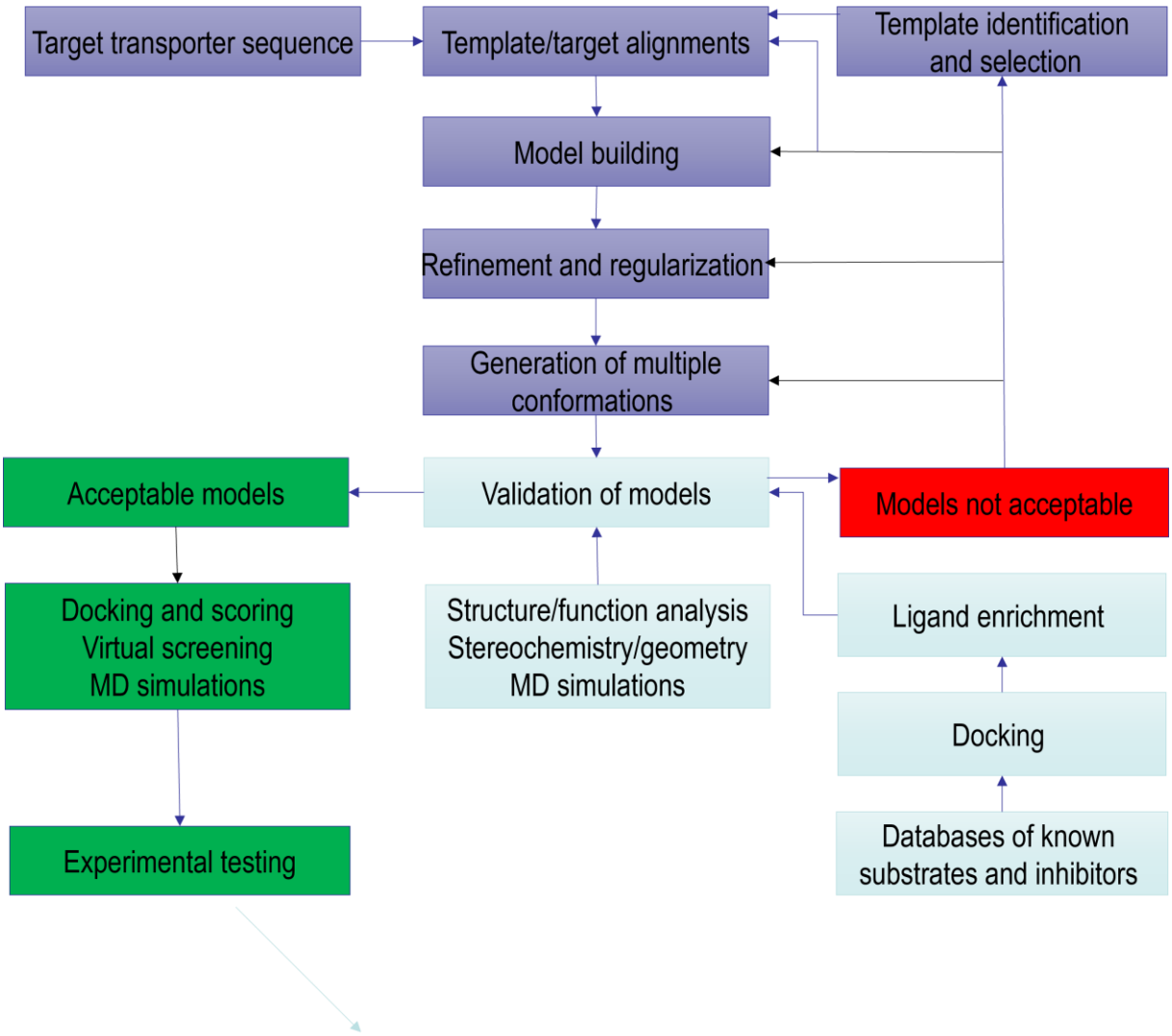
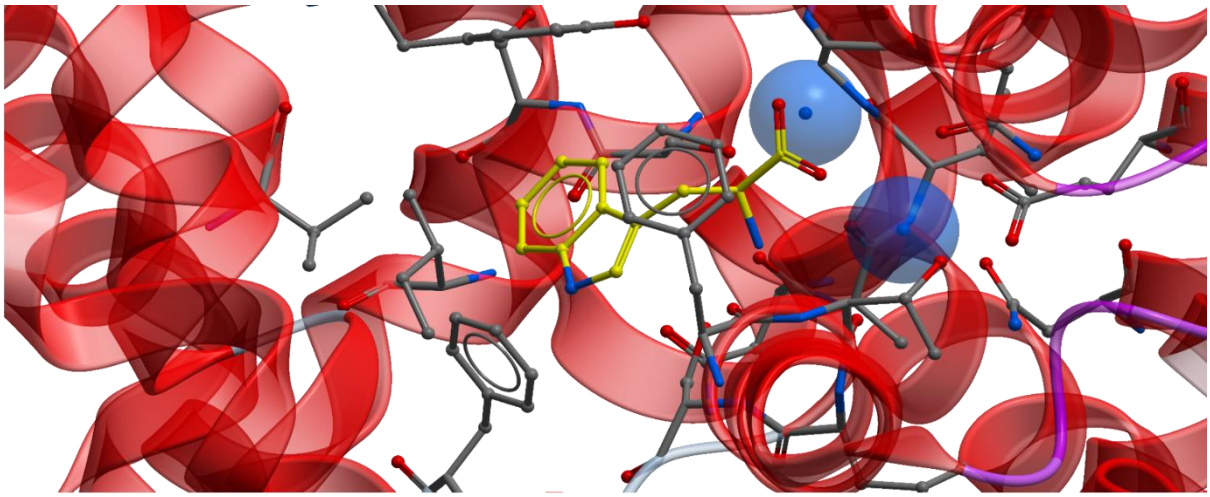
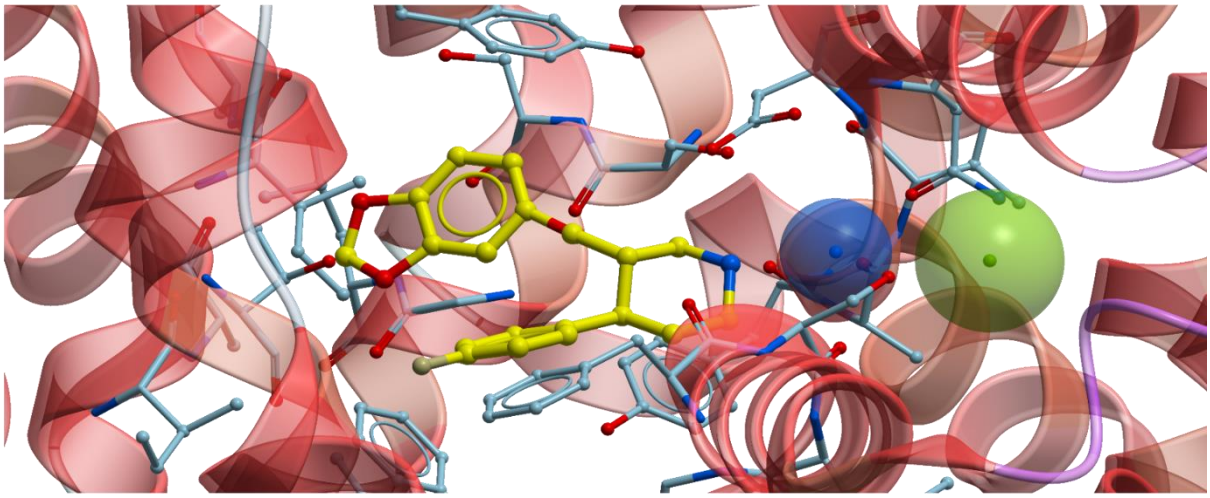

Target transporter sequence
Template/target alignments
Template identification and selection
Model building
Refinement and regularization
Generation of multiple conformations
Acceptable models
Validation of models
Models not acceptable
Docking and scoring
Virtual screening
MD simulations
Structure/function analysis
Stereochemistry/geometry
MD simulations
Ligand enrichment
Docking
Experimental testing
Databases of known substrates and inhibitors

Fig. 1



LeuT, PDB id: 3F3A



SERT, PDB id: 5I6X

Fig. 2