

MOLECULAR MODELING OF 3D STRUCTURE OF THE OXYTOCIN RECEPTOR

Discovery of novel oxytocin receptor agonists via molecular docking studies.

Gloria Antobreh

Master thesis in Pharmacy- May 2015



Acknowledgements

This study was carried out in 2014/2015 at the Department of Medical Biology, Medical Pharmacology and Toxicology, University of Tromsø,

I dedicate this thesis to the loving memory of my brother, Francis Adusei Antobreh, who passed away while the thesis was being written. I think of you everyday. Thanks for being an inspiration to me. Through the hard times your memories continued to give me needed strength to carry on.

I would like to express my deep appreciation to my supervisor, Professor Aina W. Ravna, for not only suggesting the topic for this thesis, but also for her encouragement, guidance and insightful ideas. I wish also to thank her, for providing me with a laptop for my lab work, and not forgetting the hot beverages and snacks during the “supervisor’s hours”. It has been really nice working with you.

I would also like to thank my colleagues Joseph Azumah and Sayda Colnoe for being dedicated study mates for 5 wonderful years of pharmacy school. Thanks for the friendship, support and lively working environment.

Finally, I am deeply grateful to my wonderful family; my dad, Dr, Andrew Antobreh, for encouraging me to further my education, my mom, Mrs Cecilia Antobreh for always being supportive with prayers and advice, and my sisters, Andrea Antobreh and Alfreda Antobreh, for always reminding me that I will surely make it.

Tromsø, 2015

Gloria Antobreh

Abstract

Oxytocin has been shown to be implicated in psychiatric diseases such as depression, anxiety disorders, autism post-traumatic stress disorder, and schizophrenia. As a result, oxytocin can be used as a potential treatment for these brain disorders.

However, oxytocin is a large peptide, and is therefore unable to cross the blood brain barrier. Thus, the development of new small non-peptide drugs would be of great benefit in the treatment of these neurological disorders.

In this study, new non-peptide agonists have been proposed based on homology modeling and virtual ligand screening. There is no available experimentally solved structure of the oxytocin receptor; hence three models are constructed and refined using known 3D crystal structures of evolutionary related proteins (PDB: 2Y00, PDB: 4BVN and PDB: 4LDE). The ability of the three models to discriminate between true ligands and decoys was tested and analyzed using the Receiver operating characteristics (ROC) curve.

A virtual ligand screening procedure was applied on the most suitable of the 3 models (4LDE-based model) in order to identify potential binding compounds that can be used as oxytocin receptor agonists.

The results obtained from this study are 15 compounds, which can be tested in vivo and eventually used as potential drug candidates.

Table of Contents

Acknowledgements	I
Abstract	III
Abbreviations	VII
1 Introduction	1
1.1 Context of study	1
1.2 Oxytocin	1
1.3 Oxytocin receptor	2
1.3.1 Structure	2
1.3.2 Ligand binding	3
1.3.3 Cholesterol effects	5
1.4 Drug-receptor interactions	5
1.5 Homology modeling	6
1.5.1 Template selection	7
1.5.2 Sequence-structure alignment	8
1.5.3 Building the model	8
1.5.4 Model refinement	8
1.5.5 Model evaluation	8
1.6 Docking	9
1.6.1 Receiver operating curves (ROC) curves	10
1.7 Virtual screening (VS)	11
2 Aim	13
3 Methods	14
3.1 Homology modeling	14
3.1.1 Template search	14
3.1.2 Alignment	15
3.1.3 Building OXTR models	18
3.1.4 Refinement of models	18
3.1.5 Evaluation of models	18
3.2 Docking	19
3.2.1 Agonist dataset and decoys	19
3.2.2 Ligand and receptor preparation	21
3.2.3 Identifying ligand-binding pocket	21

3.2.4	Docking of agonists	21
3.2.5	Evaluation of docking.....	21
3.3	VLS.....	22
4	Results and discussion	23
4.1	Homology modeling	23
4.2	Molecular docking	30
4.3	VLS.....	38
5	Conclusion	43
6	Further studies	44
7	Reference	45

Abbreviations

AUC	Area Under the Curve
ECL	Extracellular loop
GPCR	G protein coupled receptors
ICL	Intracellular loop
ICM	Internal Coordinate Mechanics
OT	Oxytocin
OXTR	Oxytocin receptor
PDB	Protein Data Bank
ROC	Receiver Operating Characteristic
TM	Transmembrane
VS	Virtual Screening
SBVLS	Structure-Based Virtual Ligand Screening
LBVLS	Ligand-Based Virtual Ligand Screening
OSAR	Quantitative Structure Activity Relationship

1 Introduction

1.1 Context of study

New evidence has proved that activation of the oxytocin receptor (OXTR) is likely to have positive impact on social cognition and social behavior (1). Clinical studies show that oxytocin (OT) may play a role in a number of psychiatric diseases such as depression, anxiety disorders, autism, post-traumatic stress disorder, and schizophrenia (2). OT is considered as a potential target for therapeutic intervention for these psychiatric diseases. However, the problem that arises with using OT as a drug, is that plasma OT is unable to cross the blood brain barrier (3). The blood brain barrier prevents the absorption of parentally administered OT from entering the brain for treatment of neural disorders. Thus, there is a need for new potent, selective and efficacious OXTR agonists that may be used as potential drugs. Currently, OT is used for labor induction, facilitation of placenta delivery and facilitation of milk letdown (4, 5). Available OT agonists *Pitocin* (a synthetic OT for intravenous (i.v.) and intramuscular (i.m.) administration; King Pharmaceuticals), *Syntocinon* (a synthetic OT for i.v., i.m., and intranasal administration; Novartis) and *Duratocin* (Carbetocin) a long acting OT analogue for i.v. and i.m. administration; Ferring Pharmaceuticals) are effective for obstetrics (5). These drugs are peptide agonists of the OXTR, and are associated with various limitations such as poor bioavailability due to their high hydrophobicity. Just like OT, they are too large to be taken orally or even cross the blood brain barrier. Thus, the development of small non-peptide OT agonists targeting the central OXTR that are also orally bioavailable may have therapeutic benefit for these brain disorders.

1.2 Oxytocin

OT is a peptide hormone made up of nine amino acids; Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly (NH₂). OT has a disulphide bridge between the 1st Cys and 6th Cys residue, forming a 6 amino acid ring and a C-terminal alpha amidated tripeptide residue tail (3, 6) shown in Figure 1.

OT is produced by hypothalamic neurons in the supraoptic and paraventricular nuclei of the brain (3, 7). These neurons have axonal projection sites from the hypothalamus

to the posterior pituitary. It is released from the posterior lobe of the pituitary gland in response to stimuli such as suckling, parturition or stress into the blood stream.

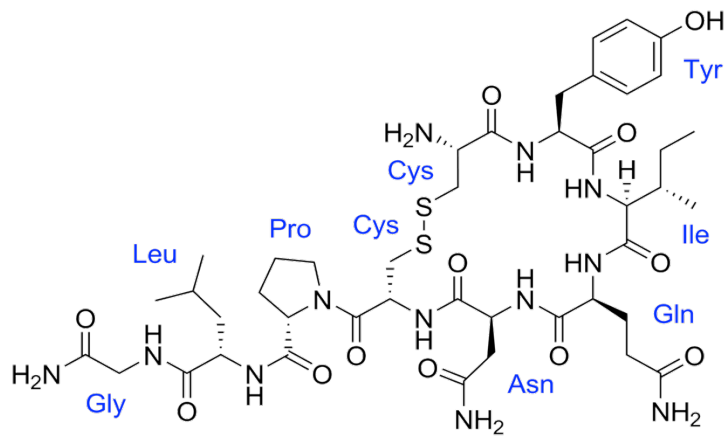


Figure 1 Structure of OT in 2D showing amino acids in blue.

Even though OT has been known as a female hormone, it is found in equivalent concentrations in both male and female neurohypophysis and plasma (4). OT is found throughout the central nervous system and the peripheral tissues. In the peripheral tissues, it is detected in organs such as uterus, placenta, amnion, corpus luteum, testis and heart, while in the central nervous system; it is detected in regions such as the cerebral cortex, hippocampus, hypothalamus, olfactory bulb and striatum. OT functions as a hormone peripherally by triggering smooth muscle contractions during childbirth, facilitating milk ejection for breastfeeding and it is involved in male ejaculation, while centrally it acts as a neurotransmitter involved in social behavior, maternal care and anxiety (3).

1.3 Oxytocin receptor

OXTRs are found in the peripheral regions such as kidney, ovary, testis, thymus, heart vascular endothelium, osteoclasts, myoblasts, pancreatic islet cells, adipocytes and several types of cancer cells. OXTRs are also widespread throughout the mammalian central nervous system (3).

1.3.1 Structure

The OXTR is made up of 389 amino acids and belongs to the G-protein coupled receptors (GPCR) class A family (see Figure 2). They consist of 7 highly conserved hydrophobic transmembrane alpha helices (7TM) spanning the membrane. The 7TM domain is connected by 3 intracellular loops (ICL1, ICL2, ICL3) and 3 extracellular

loops (ECL1, ECL2, ECL3), an extracellular N-terminal domain, and a cytoplasmic C-terminal domain (8). The helices of the GPCR contain conserved residues, which are involved in common mechanism for activation or signal transduction (3).

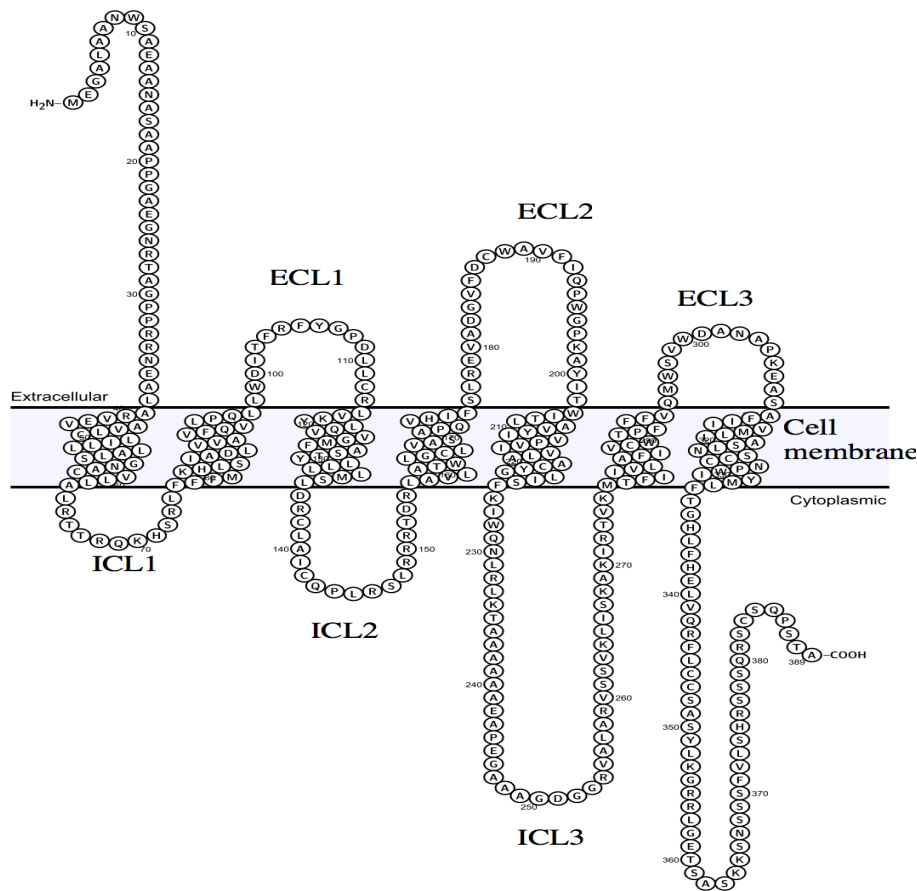


Figure 2 Residue-based diagram of oxytocin receptor generated from protter (<http://wlab.ethz.ch/protter/start/>) using uniprot id P30559.

1.3.2 Ligand binding

The extracellular regions of GPCR are responsible for ligand binding while the intracellular regions are involved in G-protein binding (ICL2, ICL3 and cytosolic extensions of helices 3 and 6) (4) .

Previous site-directed mutagenesis studies, photo affinity labeling and molecular modeling studies have demonstrated the linear C-terminal part of OT binds closer to the surface and interacts with TM2 and TM3 and ECL1, while the cyclic part is lodged in the upper one-third part of the receptor binding pockets and interacts with TM3, TM4 and TM6 (see Figure 3) (6). There are 4 residues in OXTR that studies have suggested to be important ligand binding and selectivity; Arg34, Phe103, Tyr 209 and Phe284 ((9) and ref. cited therein).The Arg34 within the N-terminal domain

is essential for OT binding. Mutational studies reveal substituting the Arg34 with Ala in the OXTR, results in a receptor with decreased affinity for OT (10).

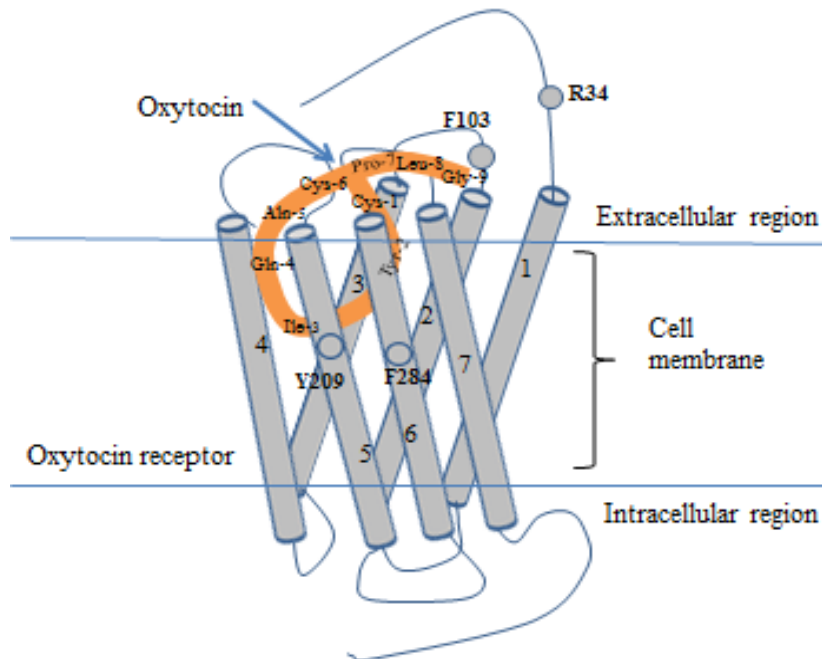


Figure 3 Model of the structure of OXTR and its interaction with OT.

However, it is not yet determined if the importance of Arg34 is due to its direct interactions with OT or whether its involvement in intra-receptor interactions necessary for a structural conformations that favor agonist binding (10, 11). Mutagenesis studies indicate Phe103 in the ECL1 is important for OT selectivity and is thought to interact with the residue at the 8th position of a peptide ligand (OT); Leu (6). Phe284 and Tyr209 of the OXTR interact with the residue at the 2nd position (Tyr) and the 3rd (Ile) position of the OT, and play a role in mediating ligand receptor interactions(9, 12). Subjecting Tyr209 and Phe284 to site-directed mutagenesis studies revealed no alteration in ligand binding properties. However, the authors concluded they might play a role in receptor activation due to the change in intrinsic activity. Ligand binding to OXTR is predicted to occur in a pocket formed by the ring-like arrangement of the seven trans-membrane domains (8).

1.3.3 Cholesterol effects

Cholesterol affects the ligand binding affinity, receptor signaling and stability of the OXTR. Cholesterol acts as an allosteric modulator and stabilizes the high affinity state of OXTR for both agonists and antagonists (3). The presence of cholesterol in the plasma membrane leads to changes in the membrane fluidity. Studies show cells with cholesterol-rich plasma membrane contain high affinity state OXTR and low cholesterol plasma membrane contains low affinity state OXTR. However, other studies have also proven that the effects of cholesterol on the OXTR are not merely due to the changes it causes in the membrane fluidity, but it is also proposed to bind to the receptor itself (13). Direct binding effects of cholesterol to the OXTR cause the receptor to adopt a more compact conformation. This effect increases the stability of the OTXR against thermal de-naturation (14).

1.4 Drug-receptor interactions

Some drugs are capable of interacting with the binding site of its receptor through covalent interactions. These are very strong irreversible forms of interactions, with bond strength of 200-400kJmol⁻¹. They occur when two atoms share a pair of electrons. An example of covalent drug- receptor binding interactions is the Phenoxybenzamine and the adrenergic receptors. However, the interactions that normally occur between a drug and the binding site of its receptor are weaker reversible non-bond interactions called intermolecular interactions. These include ionic interactions, dipole-dipole interactions, dipole-ion interactions, hydrogen bond interactions and Van der Waals interactions outlined below.

- Ionic interactions: This is an electrostatic interaction between opposite charged groups as described by Coulombs law. The interaction is influenced by water and is also pH dependent. This is the initial force that drives a drug into the binding site of a receptor. This interaction is weaker than covalent bond but the strongest of the intermolecular interactions, with bond strength of 20-40kJmol⁻¹.
- Dipole-dipole interactions: As a result of different electronegativity of some atoms, molecules can have uneven charge distribution in some regions. Dipole-dipole interactions occur when there is an attraction between the

positive region of one molecule (dipole) and the negative region of a second molecule.

- Dipole-ion interactions: These interactions occur when an ion of an opposite charge interacts with a partial charge of a dipole.
- Hydrogen bond interactions: these bonds occur when two electronegative atoms, eg, nitrogen and oxygen, interact with the same hydrogen. Normally, the hydrogen is covalently attached to one of the atoms involved, which is the donor atom, but interacts electrostatically with the other atom, which is then an acceptor. Hydrogen bonds are angle-dependent (180°) and can be influenced by water.
- Van der Waals interactions: These interactions are driven by induced electrical interactions formed between nonpolar groups/hydrophobic groups, and are weaker electrostatic interactions. They are independent of direction but are distance-dependent.

1.5 Homology modeling

The function of a protein is dependent on its 3-dimensional structure. Knowledge of the 3-dimensional structure of a protein is required to understand its ligand binding specificities. It is possible to predict amino acids sequences of proteins from DNA, and from this, their primary and secondary structures can be generated. However, the correct prediction of 3-dimensional structures of proteins has proved to be very challenging. Consequently, the structures of proteins have to be experimentally determined. The two most common ways of experimentally solving 3-dimensional structure of proteins are NMR and X-ray crystallography. However, the use of these methods to determine membrane proteins has been challenging, especially for membrane proteins. In addition, they are time consuming and labor intensive. An alternative method is homology modeling, which is a computational way of determining proteins with unknown 3-dimensional structures (steps involved are shown in Figure 4). This method is based on the fact that homologue proteins share a common evolutionary origin and therefore usually have similar 3-dimensional structures (15, 16). Thus, if one protein structure of a distinct class (eg. GPCR class A) is experimentally determined; models of all proteins in that class can be predicated based on the experimentally solved structure. Homology modeling involves

constructing a target protein (unknown 3-dimensional structure) using a template protein (known 3-dimensional structure) as a guide.

The accuracy of the model is dependent on the template selected, sequence identity and conduction of the sequence alignments between the templates and the targets as well as the quality of the templates (resolution). The following paragraph summarizes the steps involved in homology modeling approach.

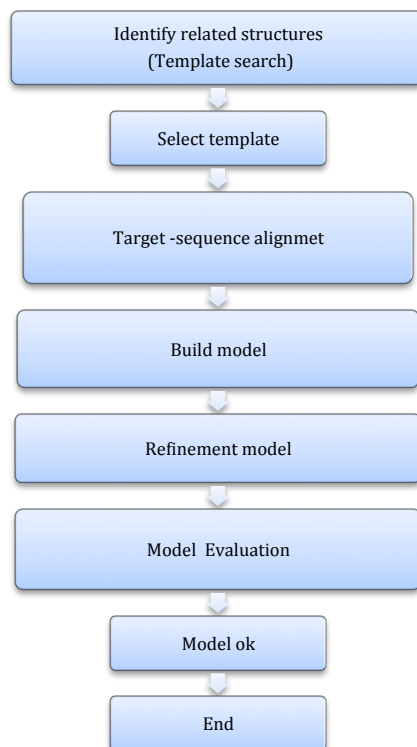


Figure 4 steps involved in homology modeling

1.5.1 Template selection

The first step in homology modeling is to find the most suitable template. In this step, the amino acid sequence of the target protein is compared to sequences of proteins with known 3-dimensional structures stored in the protein database (PDB). PDB contains a large number of proteins solved by X-ray or NMR studies. The structures of receptors are unbound or complexed with different types of ligands (inverse agonists, (ant)agonists and partial agonists). These crystal structures provide information about the different conformations of the receptors. Depending on the desired conformation of the receptor intended to be modeled (active or inactive), the most appropriate template is selected. In general, crystal structures with the high

homology to the target receptor and low resolution (as close to 1 as possible) serve as suitable templates.

1.5.2 Sequence-structure alignment

After selection of the template(s), the residues of the target sequence corresponding to the residues of the template structure are aligned to give the sequence identity. The alignment may contain deletions and/or insertion in areas where the structures of the template and target are different (low sequence similarity in some regions eg. turns). Generated alignments are carefully inspected and adjusted to improve the alignment.

1.5.3 Building the model

A 3-dimensional model of a target protein is generated on the basis of the sequence structure alignment. In this study, the ICM homology modeling macro module was used to build the models. ICM homology modeling is an automated method that generates the backbone of the models based on the aligned part of the template. Side chain conformations for residues identical to the template are then inserted. Loops are added based on conformational database searches eg. PDB with matching loop ends and upon insertion into the model non-identical side chains are assigned the most likely rotamer and optimized by torsional scan and minimization.

1.5.4 Model refinement

The models are then refined through an iterative process so that amino acids that are in too close contact with each other are adjusted and sterical clashes are avoided. The calculated energy should be as low as possible.

1.5.5 Model evaluation

Errors may occur in every model, even with low energy. The errors in models may be due to wrong side chain packing, main chain shifts, errors in unaligned loop regions, misalignment and wrong choice of template templates ((17)and ref. cited therein).

As a result, the quality of the models is evaluated to identify these errors. The evaluation checks if the bond angles, torsion angles and bond lengths are within the accepted normal ranges. The stereo-chemical properties of the models are also analyzed by comparing them to real structures with high quality.

1.6 Docking

Molecular docking has increasingly become an important tool for drug discovery (18). This approach can be used to model interactions between a small molecule and a protein at atomic level with the aim of determining the binding mode of the small molecules within the target-binding site. There are 2 basic steps in docking: ligand conformation sampling, which includes its positioning and orientation within these binding sites (pose), and ranking of the conformations through scoring functions.

There is a vast number of possible binding modes between proteins and ligands. Computers would take a long time to sample all these different binding modes. Therefore, various sampling algorithms have been developed in molecular modeling software in order to reduce the time needed to generate all possible conformations. Sampling algorithms should ideally be able to reproduce the experimental binding mode.

The aim of the scoring functions is to mark out the correct poses (true binding ligands) from the incorrect poses (decoys) in a reasonable computer time. The scoring function is not a measure of the affinity of the ligand to the protein but rather an estimation. Scoring function could be forcefield-based, empirical- and knowledge based(18).

In this study, the ICM molecular docking module version 3,8 (www.molsoft.com) was used to test the validity and predictability of the homology models and was also used in Virtual ligand screening (VLS) of compounds from publicly available databases. The Monte Carlo (19) procedure is the sampling algorithm used by ICM. This method generates ligand poses through bond rotation, rigid body translation or rotation in the space of grid potential maps calculated for the protein pocket. The obtained conformation by this transformation is tested with an energy-based criterion, which is saved and further modified to generate the next conformation, if approved. This procedure occurs in multiple steps and is iterated until the maximum level of steps is attained. The number of rotatable bonds in the ligand multiplied by the user thoroughness determines the maximum number of steps. With the help of the pre-calculated grid maps responsible for hydrogen bonding potential, van de Waals

potential, electrostatic potential and hydrophobic potential calculation, the time of the global optimization procedure is reduced. ICM uses a forcefield-based scoring function for the docking, which gives a good approximation of the binding free energy between a ligand and the protein. The ligand binding modes are scored based on the quality of the ligand protein complex.

There are different methodologies used for docking. These include the rigid ligand and rigid receptor docking, flexible ligand and rigid receptor docking, and flexible ligand and flexible receptor docking (18). The definition lies in their names, either the receptor and/or ligand is kept rigid and/or flexible during the docking process. The rigid ligand and rigid receptor is useful when an important interaction between the ligand and the receptor is already known before the docking. Using the flexible ligand flexible receptor is best for systems with induced fit paradigms but may take a lot of computing time. Flexible ligand and rigid receptor docking method is a trade-off between accuracy and computational time, and was therefore used in this study.

1.6.1 Receiver operating curves (ROC) curves

ROC curves were used to evaluate the overall predictability of the homology models. A ROC curve (Figure 5) is a graphical illustration that shows the performance and performance trade-off of a classification model. It is created by plotting true positive rate against false positive rate and the area under the curve (AUC) is calculated as an effective measure of accuracy (20).

$$\text{True positive rate} = \frac{TP}{TP + FN}$$

$$\text{False positive rate} = \frac{FP}{TN + FP}$$

where TP stands for true positive (true binding ligands classified as positives), FN stands for false negatives (true binding ligands classified as negatives), TN stands for true negative (decoys classified as negative) and FP stands for false positive (decoys classified as positives). A diagonal ROC curve from the bottom left corner to the right corner indicates that a model is a random classifier and is therefore not able to discriminate between true positives and false positives. Thus, it produces as many false positive responses as it does true positive response.

A curve closer to the to the left-hand corner and the top corner of the ROC space indicates accurate the models. Thus, these models are able to classify true positives as binding ligands and false positives as decoys. The ROC value of an excellent classifier is greater than 0.9 and the ROC value of a random classifier is 0.5.

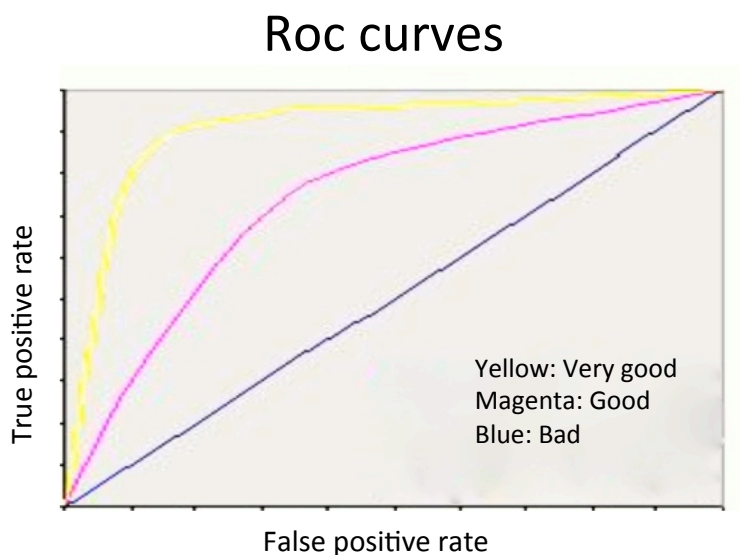


Figure 5 Graphical representation of a ROC curve

1.7 Virtual screening (VS)

Virtual screening refers to an in-silico method used to search large compound databases in order to predict their binding to a target receptor. VS does not consume valuable substance material. These compounds are virtual compounds therefore for their testing, all experimental limitations (eg. poor solubility) do not have to be taken into consideration. However, VS requires prior knowledge about the spatial and energetic criteria responsible for the binding of a particular ligand to target under investigation. Information about the 3D structure of the target (obtained experimentally or by homology modeling) or a rigid reference ligand with bioactive conformation mapping out the presumed target-binding site must be available (21).

Different VS strategies can be used depending on the amount and quality of available data. A structure-based virtual ligand screening (SBVLS) protocol that generally includes docking methods can be applied if the structure of the target is known. On the other hand, ligand-based virtual ligand screening (LBVLS) protocol that includes similarity and substructure searching, (QSAR), pharmacophore matching or 3D shape

matching can also be applied, if a rigid ligand binding to the target in question is known. However, if information about both the rigid ligand and target is known both protocols can be combined. The combination method was used in this study.

2 Aim

The 3D structure of the OXTR per date has not been solved. Ligands are recognized by the extracellular region of the receptor, and interact with the residues in a 3D environment of the receptor TM domain. With the aim of understanding the binding mode of OT agonist ligands to its receptor, in context of its 3D architecture, I developed a homology model of the human OXTR using three different published GPCR structures as templates.

The main aims were:

1. Build three homology models of OXTR in an active conformation. This step involves identifying possible template structures and conducting high sequence alignment.
2. Molecular docking of non-peptide OT agonists in the modeled OXTR to gain further insight of their interactions.
3. Evaluate the accuracy of the three models using ROC curves
4. Select one of the three models and perform VLS
5. Evaluate the VLS results for potential drug candidates.

3 Methods

3.1 Homology modeling

The 3D structure of the oxytocin receptor was not experimentally determined at the time of this study. ICM homology modeling module was used to generate a model for oxytocin receptor to aid in the discovery of new agonists. The first step of the methods was building OXTR models based on homologues proteins, followed by docking of in vitro tested agonists into the built models, then performing VLS on the most accurate of the 3 models.

3.1.1 Template search

The protein sequence of the human OXTR was extracted from [www. Uniprot.org](http://www.Uniprot.org). The Uniprot accession number P30559 was used as the query sequence (target) in finding possible homologous proteins with solved 3D structures (templates). A sequence similarity search was performed using the NCBI Basic Local Alignment Search Tool (BLAST) server. The database PDB was used to obtain information about the experimentally determined structures of the proteins. Protein-protein BLAST was chosen as the program algorithm to compare the OXTR sequence to template sequence in PDB database. The templates obtained from the search results were shortlisted based on their resolution, sequence similarity, query cover and whether or not it is complexed with an agonist. From the search, a list of 63 proteins with significant sequence homology to OXTR sequence was acquired and sorted according to the Expect value (e-value) in ascending order. The e-value is a parameter that describes the number of hits “expected “ to be seen by chance alone during a database search of a particular size. The closer an e-value is to zero, the more "significant" the match. Proteins with an e-value >0.0001 were excluded, resulting in 36 proteins remaining. All the GPCR in class A bound to an agonist were then selected using a query cover $>70\%$. The three most suitable receptors (shown in Table 1) with the highest sequence identity and highest resolution bound to different agonists, were chosen and used as templates for the modeling of OXTR.

Table 1 Protein-protein Blast search results for the 3 selected templates.

PDB ID	Structure Title	Query cover	Sequence identity %	Resolution Å	Primary Citation Author	PubMed ID
2Y00	TURKEY BETA1 ADRENERGIC RECEPTOR WITH STABILISING MUTATIONS AND BOUND PARTIAL AGONIST DOBUTAMINE (CRYSTAL DOB92)	73	26	2.50	Warne, A., Moukhametzianov, R., Baker, J.G., Nehme, R., Edwards, P.C., Leslie, A.G.W., Schertler, G.F.X., Tate, C.G.	21228877
4BVN	Ultra-thermostable beta1-adrenoceptor with cyanopindolol bound	73	26	2.10	Miller-Gallacher, J.L., Nehme, R., Warne, T., Edwards, P.C., Schertler, G.F.X., Leslie, A.G.W., Tate, C.G.	24663151
4LDE	Structure of beta2 adrenoceptor bound to BI167107 and an engineered nanobody	79	22	2.79	Ring, A.M., Manglik, A., Kruse, A.C., Enos, M.D., Weis, W.I., Garcia, K.C., Kobilka, B.K.	24056936

3.1.2 Alignment

The three templates selected for the modeling of the OXTR have the following PDB codes: 4N6H, 2Y00 and 4LDE. The OXTR was aligned with sequences of the template respectively, using the ICM's inbuilt alignment tool. The alignments were carefully inspected and adjusted using the Ballesteros nomenclature(22) as a guide. The nomenclature consists of the conserved amino acid residues found in all GPCR class A receptors. These residues are Asn (N) in TM1, Asp (D) in TM2, Arg (R) in TM3, Trp (W) in TM4, Pro (P) in TM5, Pro (P) in TM6 and Pro, (P) in TM7. The adjustment was to ensure that these conserved residues in the TM domains were in the same position in the template and target. All gaps in the initial alignments of the TM of the OXTR were deleted. This was to ensure that any insertions and/or deletions occurred in the loops area or turns rather than in the secondary structure elements (helix). Finally adjusted alignments (Figure 6,7 and 8) were used for modeling of the OXTR.

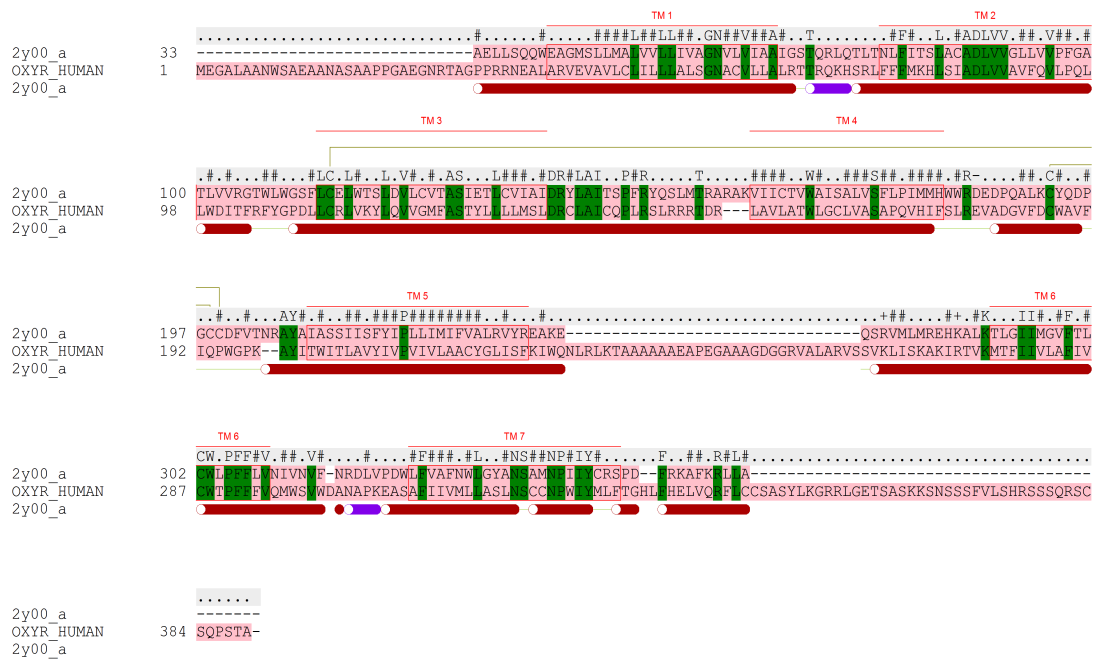


Figure 6 Sequence alignment of OXTR and 2Y00 used for homology modeling. Color code pattern: Dark green - Identical residues with template. Yellow - Disulphide bonds for the used template. The TM 1-7 are marked in the figure in red boxes. Red bars – secondary structure of the used template.

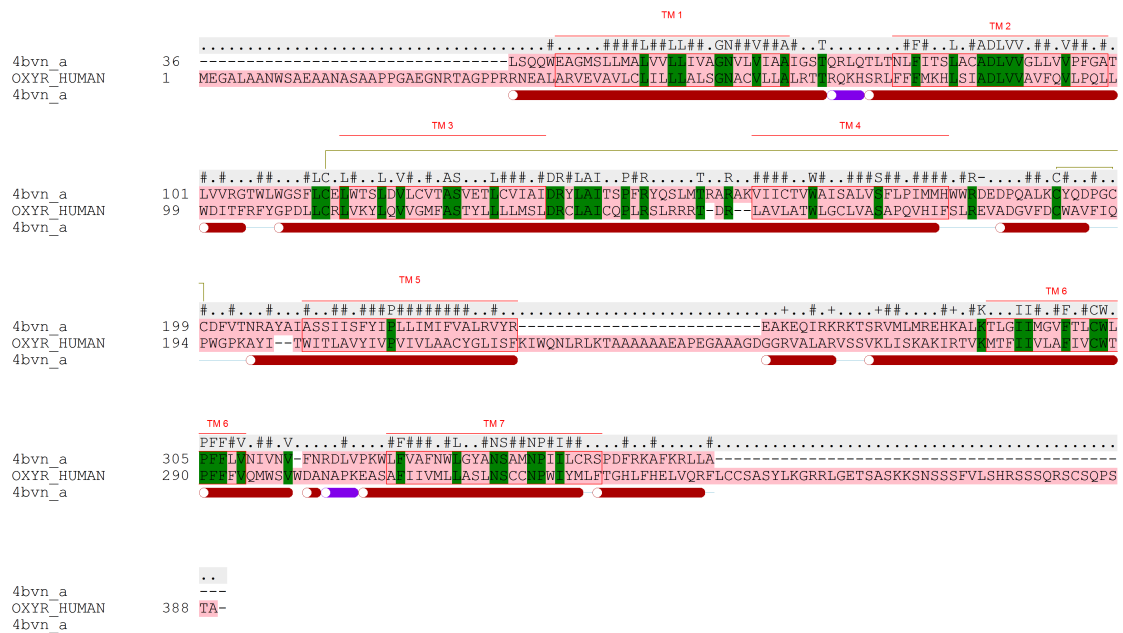


Figure 7 Sequence alignment of OXTR and 4bvn used for homology modeling. (Color code pattern: Dark green - Identical residues with template). Yellow - Disulphide bonds for the used template. The TM 1-7 are marked in the figure in red boxes. Red bars – secondary structure of the used template

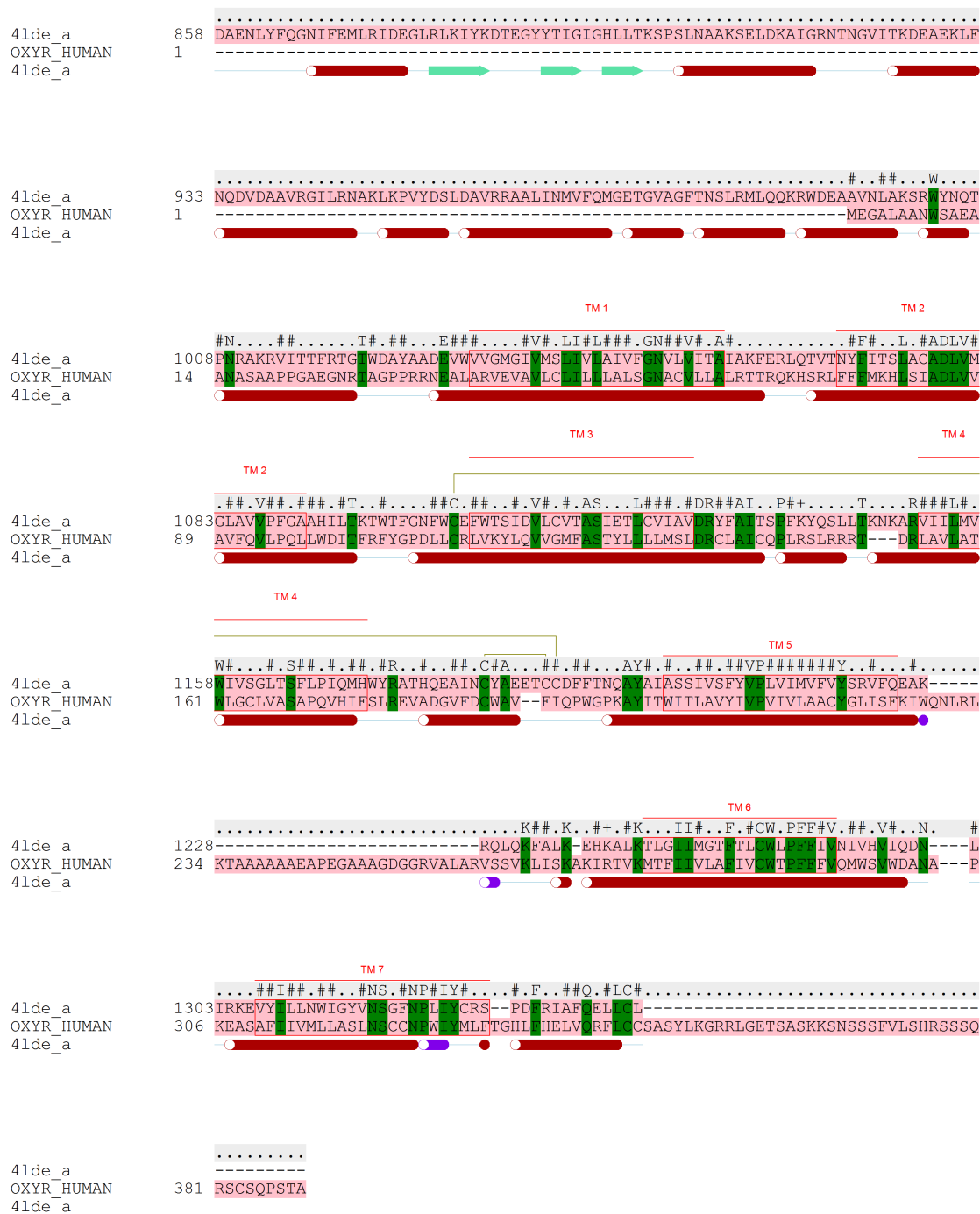


Figure 8 Sequence alignment of OXTR and 4LDE used for homology modeling. (Color code pattern: Dark green - Identical residues with template). Yellow - Disulphide bonds for the used template. The TM 1-7 are marked in the figure. Red bars – secondary structure of the used template.

3.1.3 Building OXTR models

The OXTR models were built based on the crystal structures of turkey beta-1-adrenergic receptors in complex with the partial agonist dobutamine, (2Y00(23)) with a resolution of 2,50 Å, beta-1-adrenergic receptors in complex with cyanopindolol (4BVN(24)) with a resolution of 2,1 Å and beta-2-adrenergic receptor in complex with the agonist B1167107 (4LDE(25)) with a resolution of 2,79 Å. All three structures belong to the GPCR class A receptor family and were the most suitable candidates available with regards to sequence identity, resolution and query cover.

3.1.4 Refinement of models

The ICM Refine Model macro (default settings) was used to refine the models. The macro includes 1) the program module Monte Carlo (26) fast for side chain sampling. 2) backbone iterative annealing with tethers, and 3) a second side chain sampling. Iterations of Monte Carlo-fast consist of a random move followed by local energy minimization and the iteration is either accepted or rejected based on the energy and the temperature.

3.1.5 Evaluation of models

The SAVES metasever for analyzing and validating protein structures (<http://nihserver.mbi.ucla.com/SAVES/>) was used to evaluate the models. Procheck (27), WHATCHECK (28) and ERRAT (29) were the selected programs run. Procheck provides a detailed check on the stereochemistry of a protein structure. The program compares the stereochemical parameters obtained from high resolution and wellrefined structures with a given protein structure to assess how normal or how unusual the geometry of its residues are. This program is also used to assess the quality of a protein modeled from a known crystal structure. WHATCHECK is a subset of protein verification tools from the WHATIF (30) program. This program performs widespread checking of many stereochemical parameters of the residues in the model.

Errat is a protein structure verification algorithm. It examines the statistics of non-bonded interactions between different types of atoms. The value of the error function

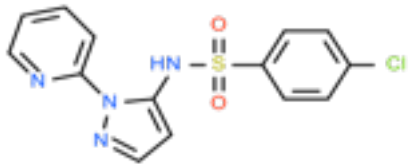
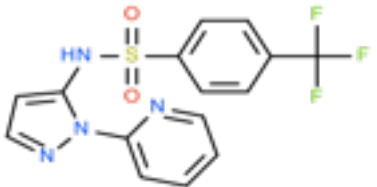
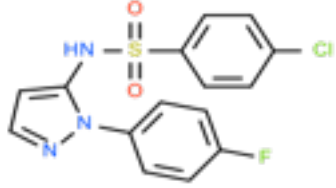
is plotted against position of a 9-residue sliding window, and the calculated by comparing statistics from highly refined structures.

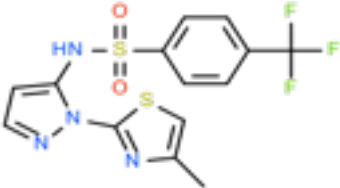
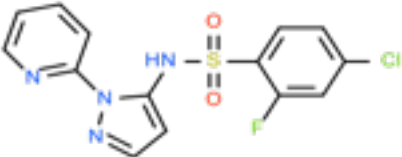
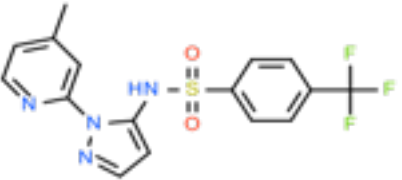
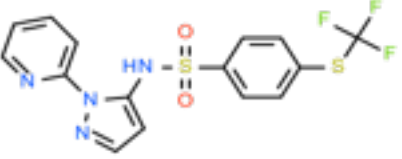
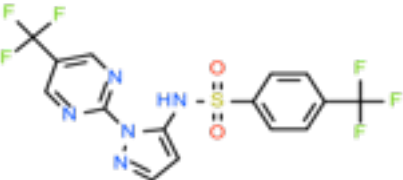
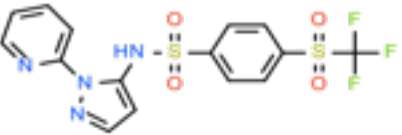
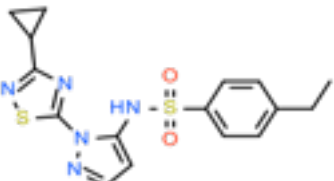
3.2 Docking

3.2.1 Agonist dataset and decoys

A set of OXTR non-peptide agonists (Table 2) was retrieved from a study by Gerard Rosse (31). These agonists belong to the compound class of pyrazolsulfonamide. Their affinities toward the oxytocin receptors tested in vitro using a calcium flux assay gave the following results shown in Table 2. Decoys were retrieved from ChEMBL using OXYR_HUMAN as a query. All compounds tested in vitro for binding to OXTR was retrieved in an sdf file, and compounds with $K_i > 10000$ nM and $EC_{50} > 100000$ nM was selected as decoys.

Table 2 OXTR agonists

Name	Structure	hEC50(μ M)
compound 1		0,019
compound 2		0,008
compound 7		2,46

compound 9		0,52
compound 14		0,019
compound 16		0,017
compound 20		0,047
compound 21		1,57
compound 23		0,753
compound 25		1,57

3.2.2 Ligand and receptor preparation

Before the docking procedure, the ligands used and the models were prepared. Both the agonists and the decoys were converted from 2D to 3D and formal charges assuming under physiological conditions (pH=7) were assigned to them where necessary, by ICM. For the models, the hydrogens were optimized, the missing side chains were hidden and HisProAsnGlnCys were also optimized.

3.2.3 Identifying ligand-binding pocket.

The right determination of the binding site of the ligand is essential for the docking process. Information about the binding sites can be obtained by comparison of the target protein with a family of proteins sharing a similar function or with proteins co-crystallized with other ligands (18, 21). The templates were used as a guide to generate the ligand-binding pocket in the models. To identify the binding pocket, the models were superimposed on their corresponding templates complexed with their agonists. Because the models and the templates have the same conformations, the space occupied by the templates agonist was empty in the models. The model residues in 5Å vicinity of the templates agonist were selected and used as the residues surrounding the binding pocket.

3.2.4 Docking of agonists

Semi- flexible docking was applied in this study. This docking process keeps the receptor rigid and the ligand is flexible. The ligand binding poses in the binding pocket is predicted by ICM with the help of Monte Carlo globalization procedure. The above-mentioned compounds were docked into the predicted binding pocket of the different models using the batch docking method of ICM. Three parallel docking runs were performed with each docking procedure and the best results were further evaluated.

3.2.5 Evaluation of docking

The overall predictability of the models was evaluated using ROC curves. The positives (agonists docked) were labeled as 1 while the false positives (decoys docked) were labeled 0. The scores from results obtained from the docking were analyzed using a ROC curve commando incorporated into ICM. The results displayed a ROC curve, and the AUC was calculated and interpreted.

3.3 VLS

The best oxytocin model, which was able to discriminate between the true positives and the false positives according to the ROC results, was used for VLS. The agonist with the highest affinity (compound 2) out of the 10 previously mentioned compounds was used to screen a commercially available database for more compounds. The database used was from www.emolecules.com. From the screening, a total of 4820 compounds with a 50% structure similarity with compound 2 were retrieved, and docked into the 4LDE-based model. The docking scores of the above agonists range from -14 to -26. The threshold score was set to -26 as a result.

4 Results and discussion

4.1 Homology modeling

The knowledge of the 3D structure of OXTR is of good help to give us important information on its ligand-binding interaction. During the time of this study, the 3D model of the OXTR was not experimentally determined; homology models were therefore constructed as a representation.

The PDB ids of templates used for the models construction were 2Y00, 4BVN and 4LDE with sequence identity of 26%, 26% and 22% and resolution of 2,50Å, 2,10Å and 2,79Å, respectively. Automated sequence alignments of the models and the templates were adjusted according to the Ballesteros nomenclature to increase the accuracy of the models. After adjustment, the sequence identity was further reduced to 21% (2Y00), 18% (4BVN), 19% (4LDE). The sequence identity between template and target correlates strongly with model accuracy (32). Thus, targets are aligned more accurately with their templates and can be generally modeled with fewer errors, when the sequence conservation is high. The 3 OXTR models had low sequence identity with their corresponding templates. This was mainly because there is limited number of available structural templates for close range homology (>50% sequence identity) modeling and furthermore, even fewer for agonists bound templates to build active conformation of target protein. However, sequence similarity alone cannot determine which template will give the most accurate model. Structural information has to be taken into consideration in the decision process of template selection for homology modeling (33).

The three final models used in this study are shown in Figure 9 (left). These models all have the typical Class A GPCR structure, consisting of 7TM domain with the connecting 3 extracellular loops, 3 connecting intracellular loops, N-terminal segment and C-terminal segment. The N-terminal segment is shown in 4LDE-based model but not in 2Y00-based model (started from residue 30) and not in 4BVN based model (started from residue 33), this is because there was no template coverage for its modeling (see sequence alignment). Previous studies have demonstrated the residue

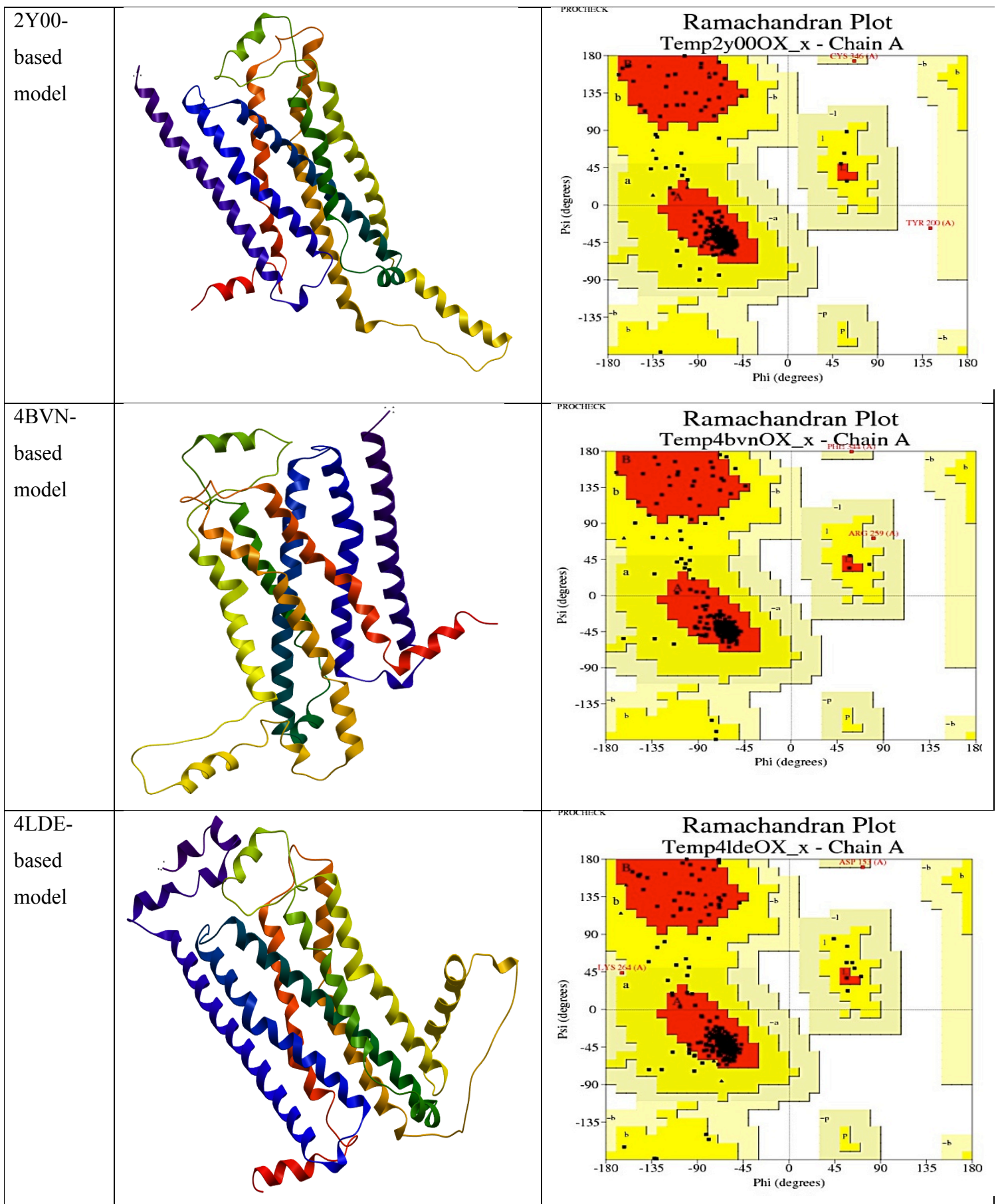


Figure 9 Left: TM helices are shown as ribbons and are spectrum color-coded, from blue (TM1) to red (TM7). The figure was generated using ICM. TM bundle are displayed with the intracellular domain at the bottom and the extracellular domain at the top. Model 1 is a 2Y00-based model, model 2 is a 4bvn-based model and model 3 and a 4LDE-based model. Right: Ramachandran plot generated via PROCHECK for Model 1, 2 and 3 respectively. PROCHECK shows residues in the most favored (red), additionally allowed (yellow), generously allowed (pale yellow) and disallowed (white color).

on the N-terminal, which might be important for ligand binding is R34 (10, 11). This residue was present in all 3 models for potential interaction. Even with low sequence identity rough models could still be built due to the common 7 TM structure of GPCR class A and also the presence of the universally conserved residues in each of the TM helices.

The right panel of Figure 9 shows a graphic display of the ramachandran plots for the corresponding models and its statistics in Table 3. The ramachandran plot, which shows the stereo-chemical evaluation of backbone psi and Phi dihedral angles of the models revealed that, generally over 90 % (except model 3), less than 9,9%, less than 0,6% and 0,3% of the residues were in most favored regions, additionally allowed regions, generously allowed regions and disallowed regions (Tyr 200), respectively. None of the residues in generously allowed regions and disallowed regions were on the located on the putative binding site of the OXTR models.

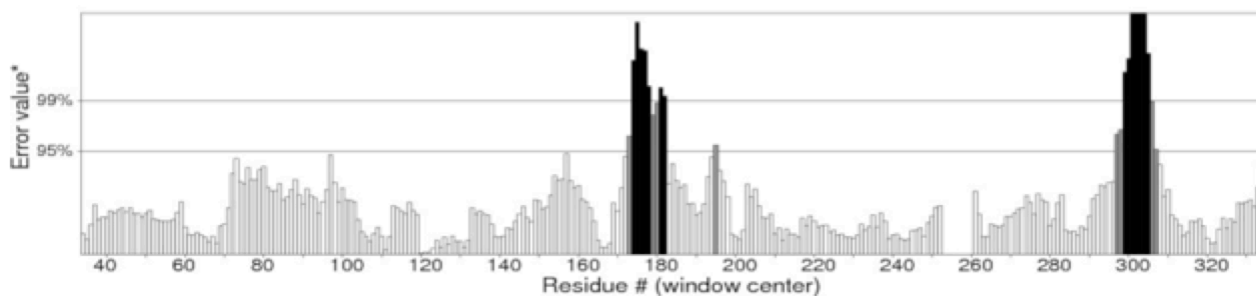
Table 3 Ramachandran plot statistics from PROCHECK

	Most favored regions	Additionally allowed regions	Generously allowed regions	Disallowed regions
2Y00-based model	92,1%	7,2%	0,3%	0,3%
4BVN-based model	90,6%	8,7%	0,7%	0,0%
4LDE-based model	89,5%	9,9%	0,6%	0,0%

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favored regions (PROCHECK). From the table we see that 2Y00-based model and 4BVN-based model are within this limits (92,1% and 90,6% respectively) while 4LDE-based model is outside of that limit but very close (89,5%). This indicated all the models where of good quality.

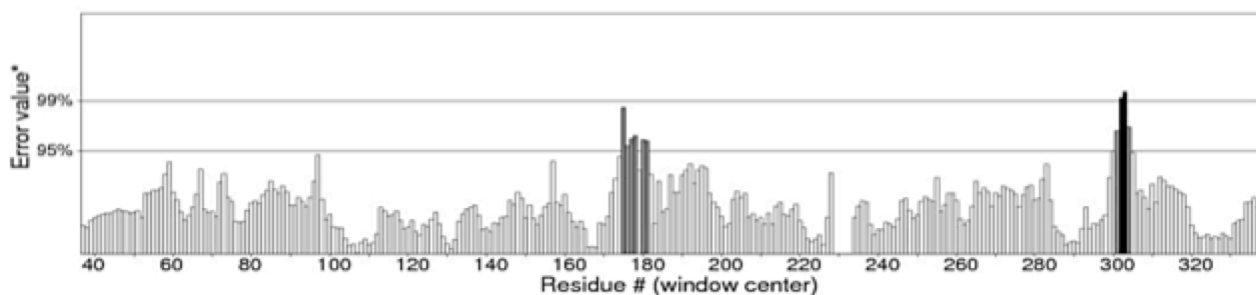
2Y00-based model

Overall quality factor**: 91.722



4BVN-based model

Overall quality factor**: 96.667



4LDE-based model

Overall quality factor**: 92.683

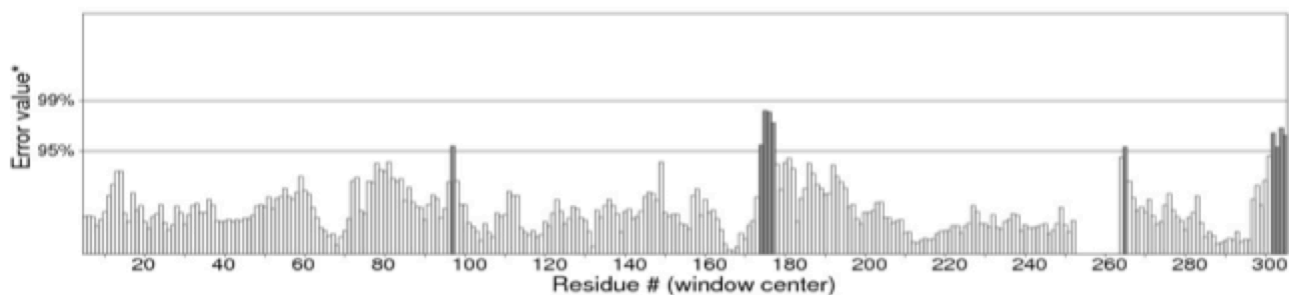


Figure 10 ERRAT plots for 2Y00-based model, 4BVN-based model and 4LDE-based model. *The two lines drawn on the error axis show the confidence with which it is possible to reject regions that exceed that error value. ** Expressed as the percentage of the protein for which the calculated error value falls below the 95% rejection limit. Good high resolution structures generally produce values around 95% or higher. For lower resolutions (2.5 to 3 Å) the average overall quality factor is around 91% (Errat).

ERRAT shows the analysis of the statistics of non-bonded interactions between different atom types, with higher scores indicating higher quality (29). 4BVN-based model had the highest ERRAT score at 96,667, while 2Y00-based model and 4LDE-based model had close scores of 91,722 and 92,683, respectively (Figure 10). The generally accepted range is >50 for high quality models (34). Therefore, the analysis revealed that the backbone conformation and non-bonded interactions of the OXTR models fits well within the range of a high quality model.

Using ICM, the models were superimposed on their corresponding templates (see Figure 11) and the values between their backbone c-alpha atoms were calculated. This was to check how well the models resemble their templates. The RMSD value describes the structural difference between the models and their templates, with low RMSD value indicating less structural difference. 4BVN-based model with a sequence identity of 26% had the lowest RMSD value at 0,260Å, followed by the 2Y00-based model with a sequence identity of 26% and RMSD value of 0,272Å and 4Lde-based model with a lower sequence identity of 22 % and an RMSD value of 0,306Å. Despite low sequence identity, the models were well superimposed on their templates, but had different conformations at certain regions. Most difference between the models and the templates were in the ICL3 in all the models. OXTR models appeared to have a longer ICL3 with different conformation compared to all 3 templates used. The template of the 4LDE-based model had a longer N-terminal segment, which explains the slightly higher RMSD value. The low RMSD values and the small variability in some regions between models and the templates show there is little structural difference between them.

2Y00-based Model

Rmsd: 0,272Å

4BVN-based Model

Rmsd: 0,260Å

4LDE-based Model

Rmsd: 0,306Å

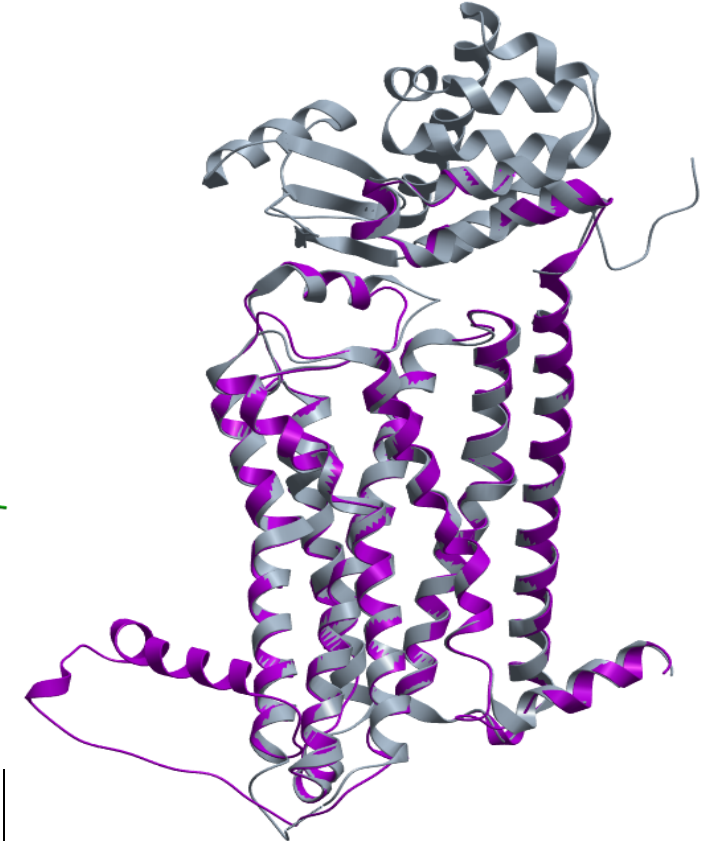
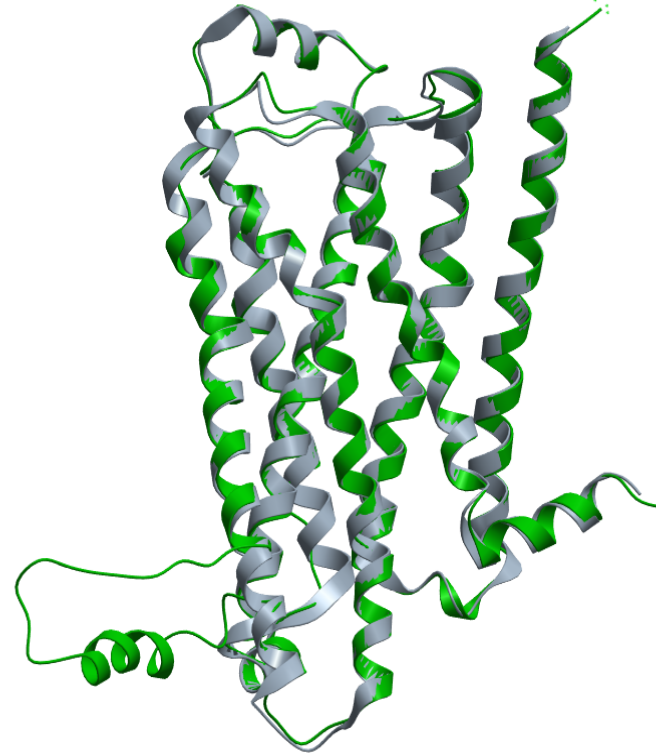
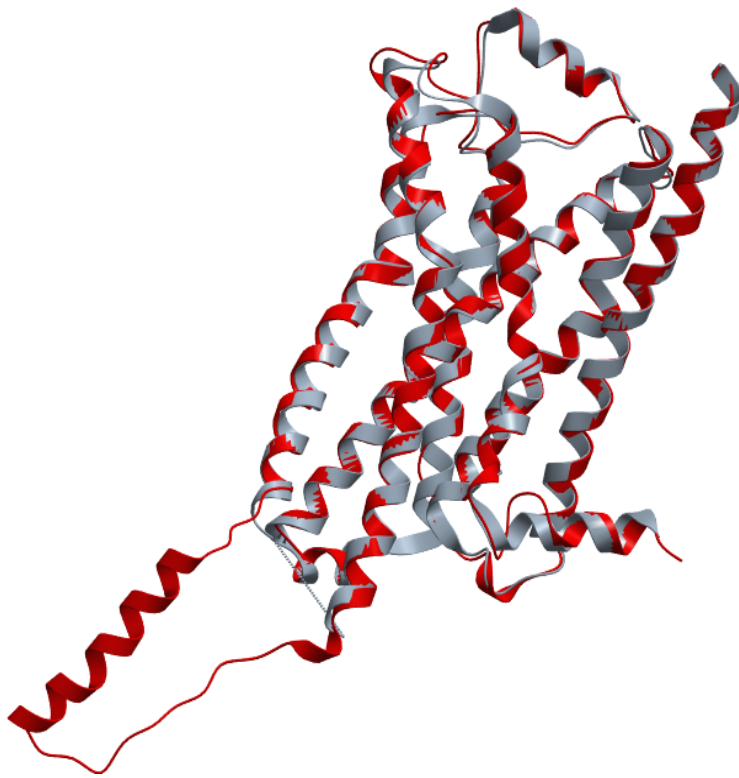


Figure 11 Superimposition of the models and their corresponding templates represented as ribbons. Left: 2Y00-based model (red), middle: 4BVN-based model (green) and right: 4LDE -based model (violet). Templates are shown in grey color.

The most outstanding difference between the 3 superimposed OXTR models was the ICL3 having different orientations (see Figure 12). The crystal structure of the template used in modeling the OXTR model did not contain the ICL3. As a result, the OXTR models were not built on structural templates. ICL3 of the 2Y00-based model (violet) was a bit separated from the other 2 models and was oriented into the intracellular part. Since the ICL3 is found in the intracellular part, it might not affect the binding interaction of the ligand, but it might affect the g-protein coupling since it is located in the intracellular part.

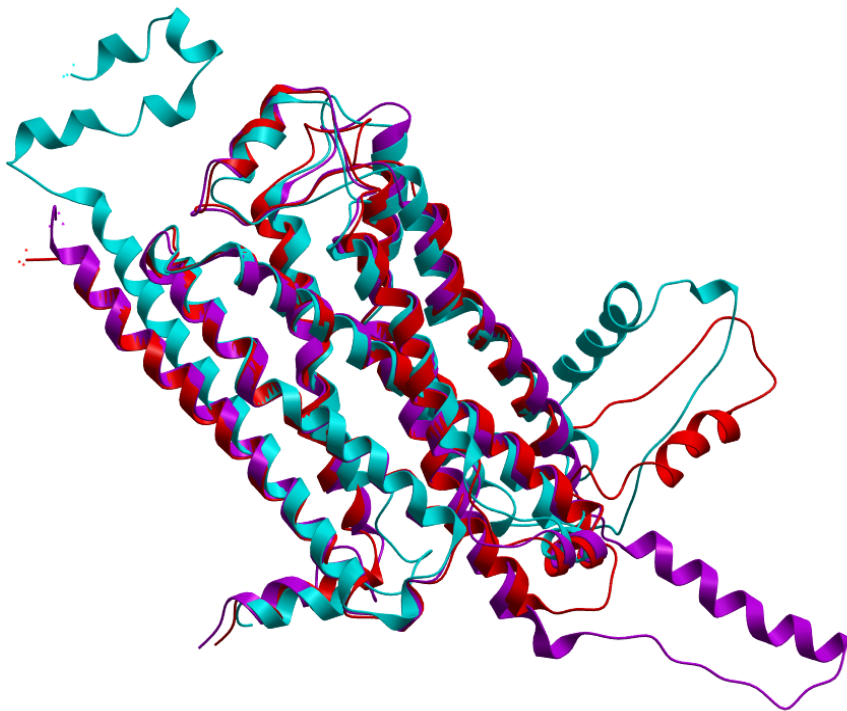


Figure 12 Superimposition of the 3 models represented as ribbons. Cyan: 4LDE- based model, red: 4BVN-based model and violet: 2Y00-based model. (Rmsd: 0,944)

4.2 Molecular docking

Molecular docking is the most widespread approach for determining protein ligand interactions (35). It is a method used to predict potential ligand binding site of a target protein. Studies on the binding sites of OXTR may provide insightful information about key interactions and characteristics that aid in the designing novel drugs. Molecular docking of known agonists into the putative binding sites of created OXTR models were carried out to explore its overall predictability and aid in designing rational agonists.

Class A GPCR share a common binding pocket located deeply within the 7TM domain for small molecule agonist and antagonists (9). This has been confirmed by several GPCR crystal structures in complex with small ligands including the templates used in this study. Since OT is much larger than the agonists used in this study, there is a likelihood that the agonists would bind in a similar manner as the agonists in the templates. Based on this hypothesis, the binding site of the template bound to a small agonist was used as a guide for mapping OXTR binding site. Gln119, Lys116 and Met123 in helix 3, Gln295 in helix 6 of OXTR putative binding site have been proposed by previous studies of molecular dynamic simulation (36) and molecular modeling (9) as residues most likely to have contact with small ligands. The 4 residues suggested were also present in the putative binding sites of all 3 OXTR models for docking the compounds (see Table 4).

Table 4 shows residues forming ligand-binding site for 2Y00-based model, 4BVN-based model and 4LDE-based model. Residues most likely to have contact with small ligands are colored in green.

	Residues forming pocket
2Y00-based model	Q96, W99, D100, K116, Q119 , V120, M123 , K198, Y200, I204, T205, V208, W288, F291, F292, Q295 , I312, M315, L316
4BVN-based model	Q92, W99, K116, Q119 , V120, M123 , K198, I201, I204, T205, V208, W288, F291: F292, Q295 , F311, M315
4LDE-based model	Q92, Q96, W99, K116, Q119 : V120, M123 , I192, I201, I204, T205, V208, W288, F291, F292, Q295 , F311, I312, M315

The residues in the putative binding sites of the OXTR models does not include the 4 previously mentioned residues important for OT binding; Arg34, Phe103, Tyr 209 and Phe284. These residues are also not among the residues most likely to have contacts with small ligands. This might be due to their location. As shown in (Figure 13), Tyr 209 and Phe284 are located deep in the binding pocket below the putative binding site, while Arg34 and Phe103 are located on the entrance in the extracellular side.



Figure 13 OXTR models superimposed showing the binding site residues in wires and residues important for OT binding in sticks and labeled. Orange: 2Y00 based model, yellow: 4BVN based model and green: 4LDE based model.

Small agonist can stabilize proposed active conformation of GPCR by acting as a molecular glue interacting with residues deep in the main binding pocket between the helices, while larger agonists and peptides can stabilize the active conformation in a similar way by acting as a velcro on the connecting loops and at the extracellular ends of the helices (37). Oxytocin is a peptide hormone and it is large enough to interact with residues in the deep binding pocket and the ECL residues in order to stabilize the proposed active conformation. On the other hand, depending on the size of the non-peptide agonists, they might not be able to

interact with all residues important for OT binding, but still stabilize the active conformation. The putative binding site was based on this hypothesis.

62 decoys and 10 active agonists were docked into the proposed binding sites of OXTR models to evaluate the ability of the models to distinguish between them. The experimental binding affinities of the 10 compounds used for the docking are shown in table 2. These compounds belong to the pyrazolsulfonamide functional group (see Figure 14). They differ from each other by the A₁, R₁ and R₂ side groups. A₁ and A₂ groups are aromatic. A₁ in the 10 compounds are benzene rings.

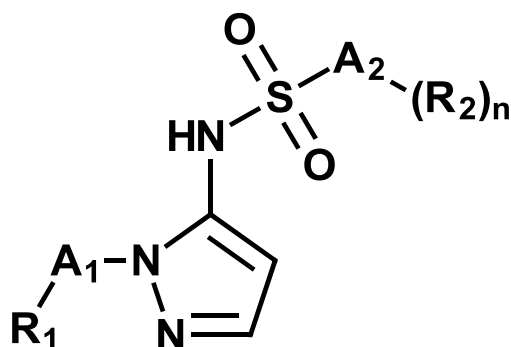


Figure 14 Pyrazolsulfonamide functional group

The binding modes of some of the docked agonists were investigated by visual inspection, in addition to analysis of the scores for the ligand-receptor complexes. The docking results show most of the agonists interact with residues closer to the surface of the proposed binding pocket in the models, rather than deeper in the pocket as anticipated and has docking scores ranging from -14 to -26 in the most predicative model (see Figure 15). The reason for this could be the binding site was too narrow in this conformation for the agonists to penetrate deeper.

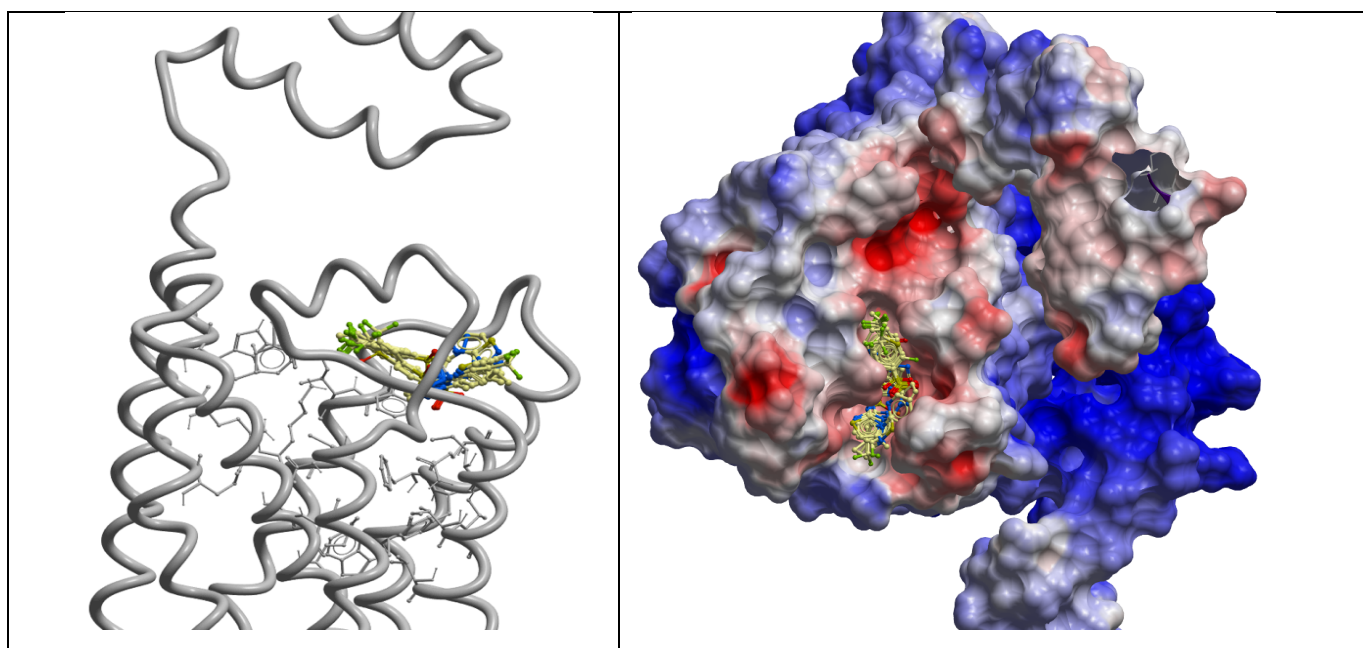


Figure 15 Binding mode of pyrazolsulfonamide Left: Residues in the binding pocket shown as wires and 4LDE-based model shown in ribbon worms in grey. Right: 4LDE-based model in electrostatic potential (from the extracellular region). Red: negative, blue: positive and white color: neutral.

The left of Figure 15 shows the binding site of the 4LDE-based model on the surface of the receptor model. All the agonists are nicely docked in the pocket. The right panel of Figure 15 shows the 4LDE-based model electrostatic potential from the extracellular side, with the 10 docked agonists in the putative binding site. The Figure shows the pocket formed on the surface of the receptor model with the agonist inside.

The proposed binding mode of compound 68 was predicated to have the lowest score at -26. However, this compound was not the most active compound determined experimentally rather the second to the best. Compound 68 differs from compound 64, which is the experimentally most active agonist by an additional methyl group linked its A₁ pyridine ring whereas compound 64 has just the plain pyridine ring without an attached group (see table 2).

Compound 64 had a slightly higher score at -24. The benzene ring connected to the trifluoromethyl and sulfonamide group on this compound is predicated to interact between Phe311 and Trp195 through stacking interactions (see Figure 16). There was a hydrogen bond observed between Gly196 and n2 of compound 64 at a distance of 2.33Å. There might also be a possible hydrogen bonding interaction between its n1 and the side chain oxygen on Ser298 (not shown in Figure 16) of the OXTR. In reality these interactions are not static as

shown, there are movement and flexibility between the molecules such that they can find each other. This picture is just a snapshot of how compound 64 might be interacting in reality.

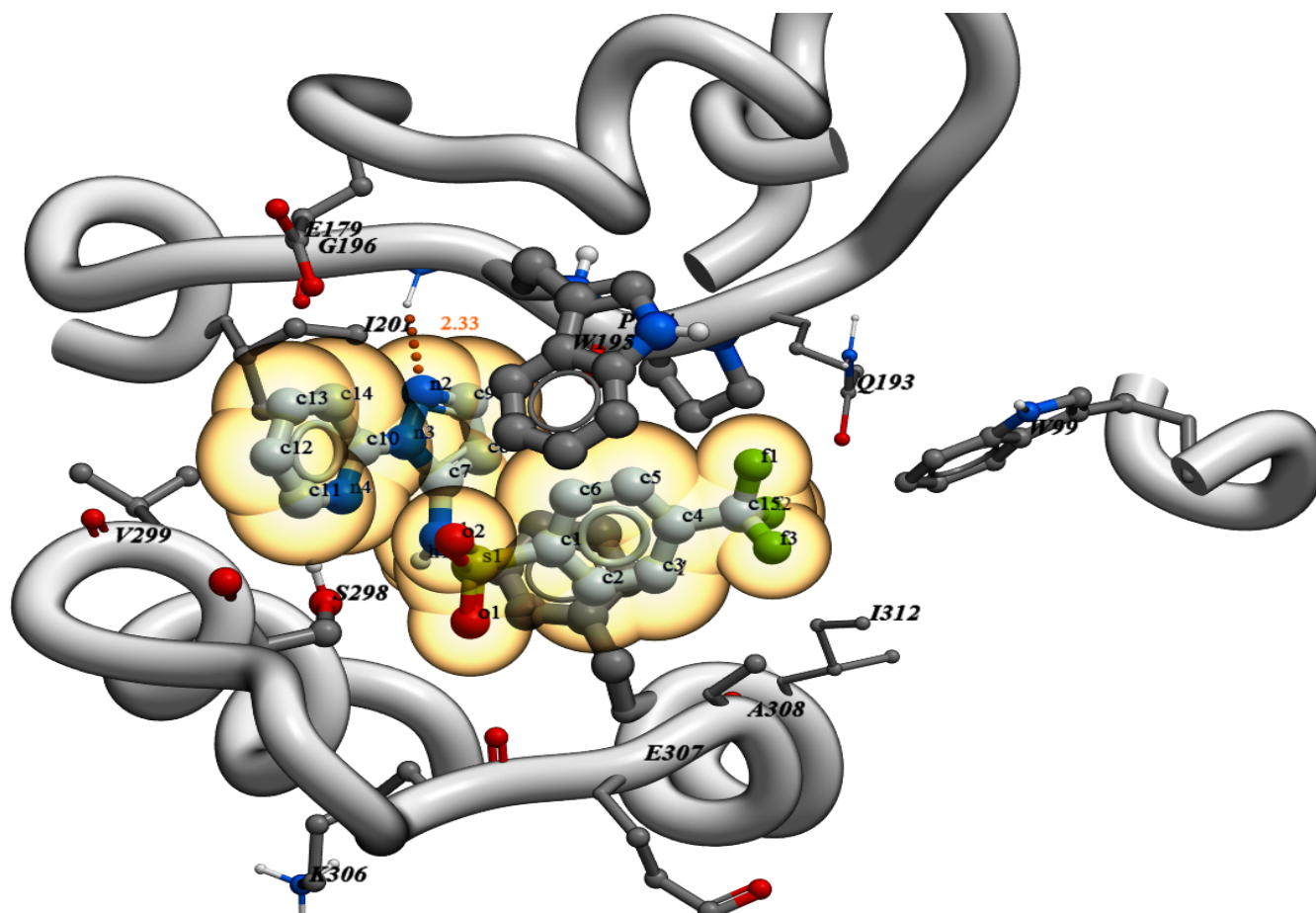


Figure 16 Predicated binding interactions between compound 64 and OXTR.

Compound 68 had similar conformation and interactions as compound 64 (see Figure 17). The difference in predicated interaction between the two compounds and the OXTR receptor model is that, the methyl group on the compound 68 pyridine ring might probably be engaged in hydrophobic interaction with Ile201 and Val299. A hydrogen bonding interaction is shown between n1 of compound 68 and Ser298 of the receptor at a distance of 2,46. A possible hydrogen bonding interaction between Gly196 and n2 of compound 68 might be present even though it is not shown on the Figure.

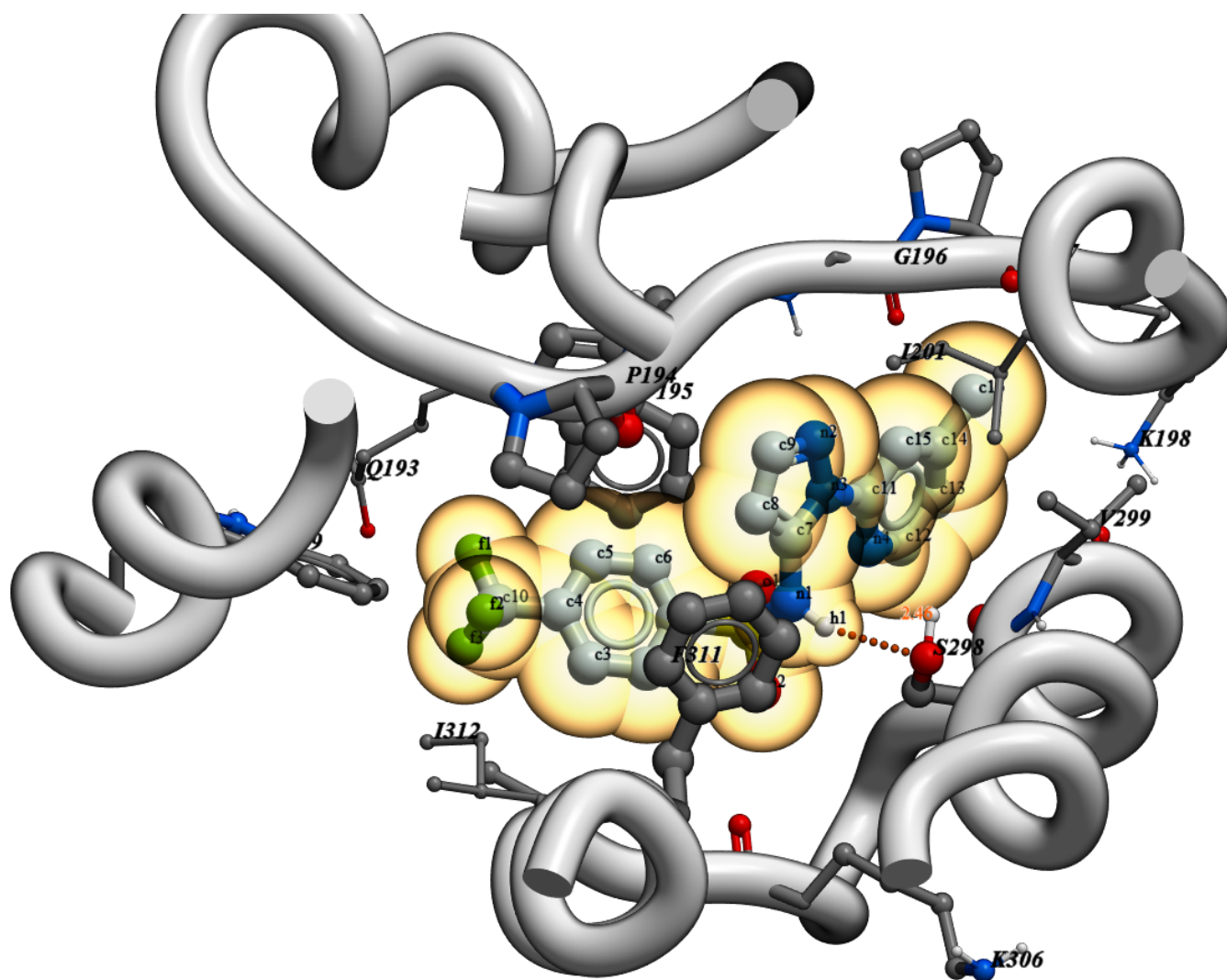


Figure 17 Predicated binding interactions between compound 64 and OXTR.

According to my findings, the residues that might be important for pyrazolsulfonamide binding are Ser298, Gly196, Phe311 and Trp195. Unfortunately, there are no mutagenesis data to confirm that these residues actually play a role in agonist binding. These residues along with other residues surrounding the putative binding sites of the OXTR models can aid in selecting residues for site directed mutagenesis studies. If experimental mutagenesis studies confirm these residues to be important as proposed, we can conclude that the models are partly correct and the residues are important.

GPCRs exist in equilibrium between active and inactive conformation. Agonist can exerts their action by stabilizing the active signaling conformation. Agonist-independent active state of the OXTR has been previously described (38). Chemically diverse agonists can bind to the active conformation, but they do not need to bind to a common residue(s) in order to exert their activity(37). This means the residues proposed to be important for OXTR agonist

activation in this study, might be specific for agonist having the similar size and chemical nature to pyrazolsulfonamides and not in general for OXTR agonists. This might also be an explanation for why the 4 above mentioned residues, found in other studies to be responsible for OXTR activation, and also the other 4 residues proposed to come close in contact with small agonist did not interact with the pyrazolsulfonamides.

Additional mutagenesis studies together with my results can help lead in the right direction but only an X-ray structure of OXTR co-crystallized with pyrazolsulfonamide can elucidate its possible binding modes.

The ROC curve was used to evaluate the ability of the OXTR models to distinguish between active compounds over decoy molecules. Figure 18 shows the ROC curves of the 3 OXTR models. The ROC curve of 2Y00-based model is closer to the diagonal line than the top left corner. Its calculated AUC value is 75. This means it is fairly good in distinguishing between actives and decoys. The ROC curve of 4BVN-based model crosses the diagonal line. Many values are on the right side of the diagonal line. This shows that this model is worse than the random classifier. Thus, it classifies more decoys as positives than the actives as positives. The calculated AUC value is 62. Compared to the other 2 models, the 4BVN-based model has the worst distinguishing properties. The ROC curve of 4LDE-based model is closest to the top left and farthest from the diagonal line and it also has the highest AUC value at 85, compared with the 2 other models. Thus, the 4LDE-based model had the best performance in distinguishing between the actives and the decoys compared to 2Y00- and 4BVN-based models. Therefore it was used to run the VLS.

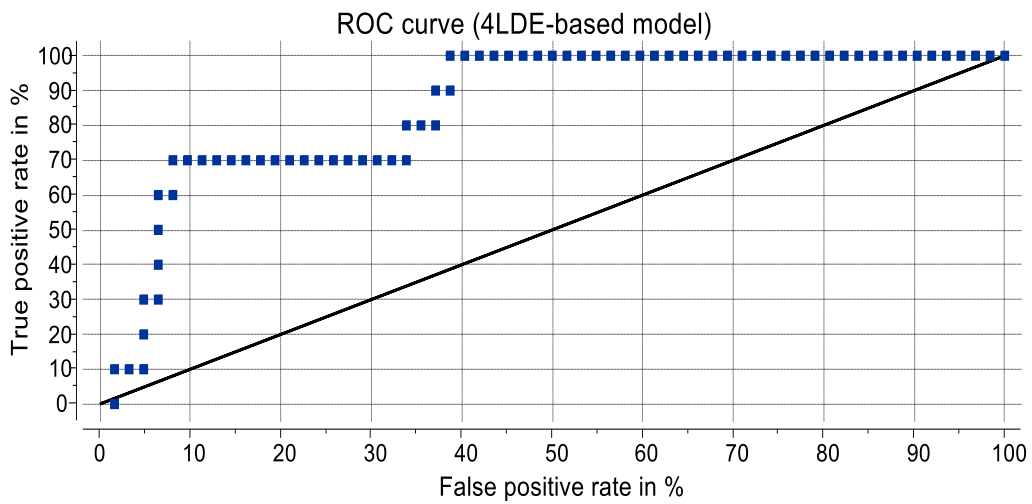
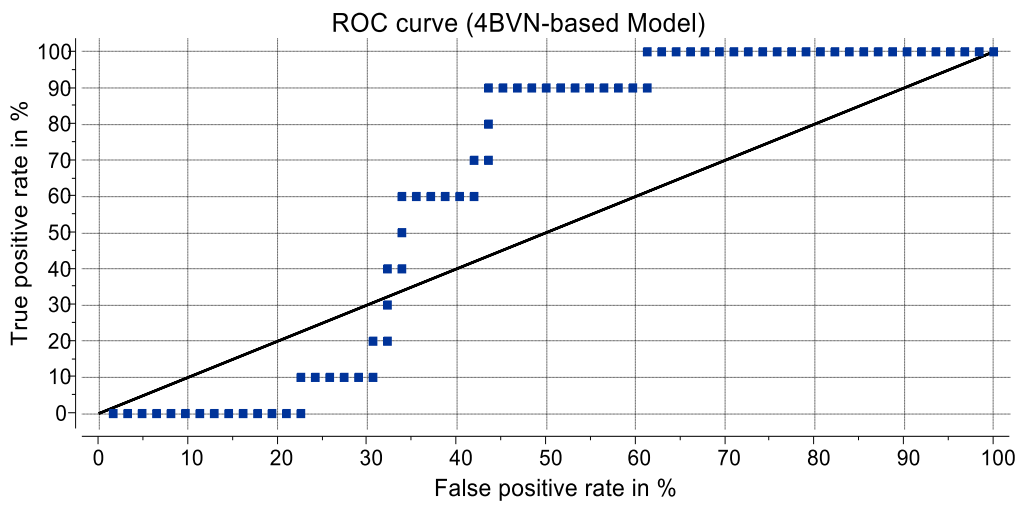
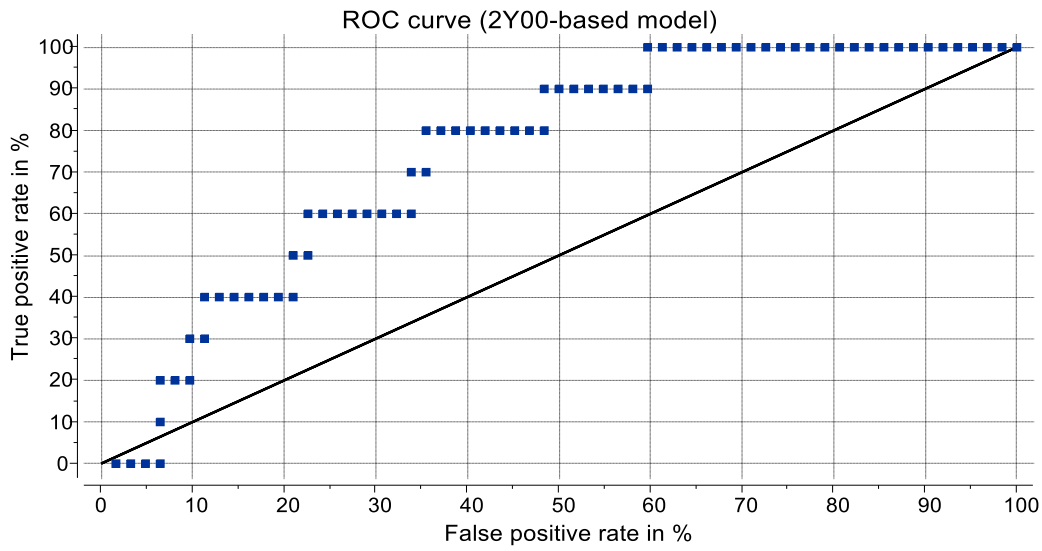


Figure 18 ROC curves of the three OXTR models

4.3 VLS

Compound 64, which was experimentally ranked as the agonist with the highest affinity was used as a candidate ligand to screen a commercially available database for potential compounds that can be used as potential drugs. From the screening, a total of 4820 compounds with a 50% structure similarity with compound 64 were retrieved, and docked into the 4Lde-based model. From the docking of the 4820 compounds, 53 compounds had scores at from -26 and better, with the best score at -30.

All the compounds interacted with residues on the surface of the putative binding site of the 4LDE-based model; the same location as where the known agonists bind.

Aromatic rings on most the compounds interacted with the Phe311 and the Trp195 of the receptor through stacking interactions and the compounds had several hydrogen bonds with the neighboring amino acids containing electronegative atoms of the OXTR models. These amino acids might be important for small ligand binding to OXTR and might help in identification of how agonists activate the receptor.

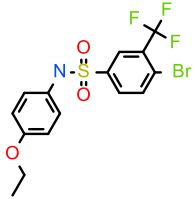
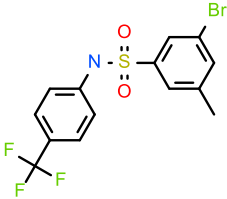
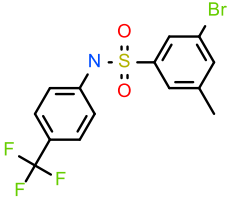
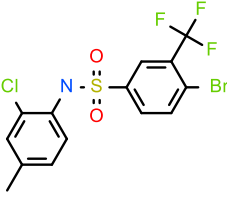
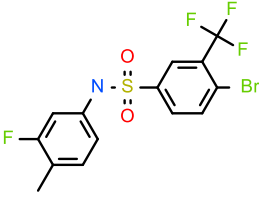
Compounds that are likely to cross the blood brain barrier usually have the following physiochemical properties(29, 39, 40);

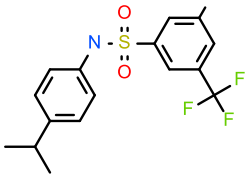
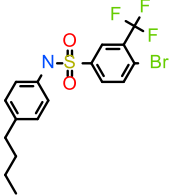
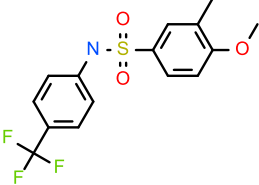
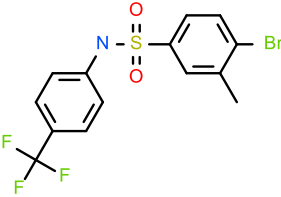
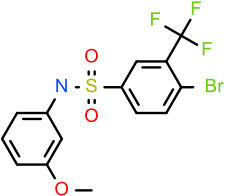
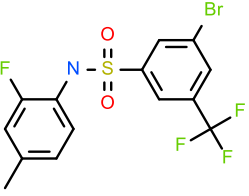
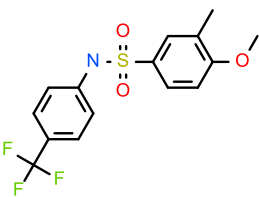
- $N + O < 6$
- $PSA < 60-70\text{\AA}$
- $MW < 450$
- High logP

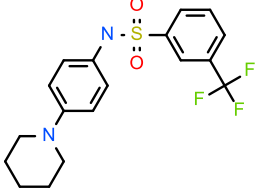
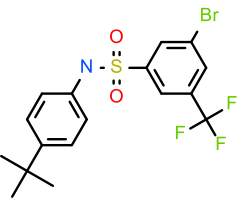
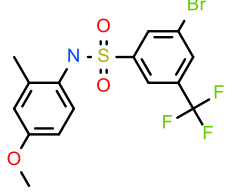
The results obtained from the VLS were shortlisted based on their score, then iterated on physiochemical properties essential for blood brain barrier crossing. The selected compounds suitable for further in vitro testing are shown in Table 5.

There is still lack of understanding in physics related to the binding process and scoring. As a result, the compound with the highest score does not necessarily mean the compound with the highest affinity. There is also no guarantee that these proposed compounds will bind to the OXTR when tested experimentally. This method is mainly used as a guide to find rational compounds for in vitro testing. It is the best option, than alternatively testing a database of compounds without prior knowledge.

Table 5 Proposed OXTR agonists selected from the virtual ligand screening.

Structure	Molecular formula	Score	LogP	PSA	HBA	HBD	Molecular Weight
	C15H13BrF3NO3S	-28,97	5,10	48,01	5	1	422,98
	C14H11BrF3NO2S	-27,61	5,05	40,89	4	1	392,96
	C14H11BrF3NO2S	-27,58	5,05	40,89	4	1	392,96
	C14H10BrClF3NO2S	-27,38	5,53	40,19	4	1	426,93
	C14H10BrF4NO2S	-27,30	5,08	40,89	4	1	410,96

	C16H15BrF3NO2S	-26,99	5,77	40,89	4	1	421,00
	C17H17BrF3NO2S	-26,87	6,40	40,89	4	1	435,12
	C15H14F3NO3S	-26,68	4,17	48,52	5	1	345,06
	C14H11BrF3NO2S	-26,64	4,93	40,89	4	1	392,96
	C14H11BrF3NO3S	-26,49	4,62	48,43	5	1	408,96
	C14H10BrF4NO2S	-26,47	5,20	40,19	4	1	410,96
	C15H14F3NO3S	-26,37	4,17	48,52	5	1	345,06

	C18H19F3N2O2S	-26,06	4,97	44,34	4	1	384,11
	C17H17BrF3NO2S	-26,05	6,27	40,89	4	1	435,01
	C15H13BrF3NO3S	-26,01	5,03	47,73	5	1	422,98

There are various limitations associated with the models presented in this study. These are summarized in the following paragraphs below (41):

The qualities of the models are dependent on the quality of the crystal structures. Any error in the crystal structure is inferred on the model. The resolution used in this model was less than 3, which is considered low enough for good quality homology modeling.

Homology modeling in combination with experimental data increases the strength of the results, due to the lack of experimental data on OT-OXTR binding; the results presented in this study might be limited. Thus, the models have to be experimentally tested and readjusted if needed.

Although the models in this study might be of good quality as described earlier, it is still a weak representation of what might be happening in living cells in reality. The models are static and do not consider factors such as dynamic changes, loop movements, allosteric interactions (eg. with cholesterol), membrane environment, solvents and other parameters inverting in living cells.

All in all, the models presented in this study are to be used as tools for experimental studies such as mutagenesis studies. Currently, it is the only available option to study ligand drug interaction whiles waiting for a crystal structure.

5 Conclusion

Three models have been constructed in order to explore the molecular interactions of the OXTR. The models constructed were of high quality to be used for docking studies. Out of the 3 models presented in this study, the 4LDE- based model has been found to be the most effective in discriminating between true agonists and decoys. This model can be used as a tool for experimental studies. The homology models created in this study has allowed us to propose residues from the putative OXTR binding site that may be involved in agonist binding interactions, making them good candidates for site-directed mutagenesis studies. From the VLS performed on 4LDE-based model, 15 compounds have been selected for further in vivo testing. The key interactions and characteristics involved of the OXTR agonist binding still remains ambiguous due to the lack of experimental data. Consequently, it has been challenging to use homology models to map the binding mode of known OXTR agonists as bases in theoretically determining new novel agonists without any reasonable doubt.

6 Further studies

As the Figure below illustrates, there is a long process involved both in the discovery and development process in finding of new drugs. The work done in this study is only in the beginning stages. 15 compounds have been suggested as potential drug candidates for the OXTR. The compounds generated have to be tested in vitro to find out their affinities to the OTXR. If the compounds bind as theoretically proposed, they can be further optimized and used for in vivo OXTR affinity testing. Molecule(s) that cross this stage can then be used as potential drug candidates for the treatment of psychiatric disorders.

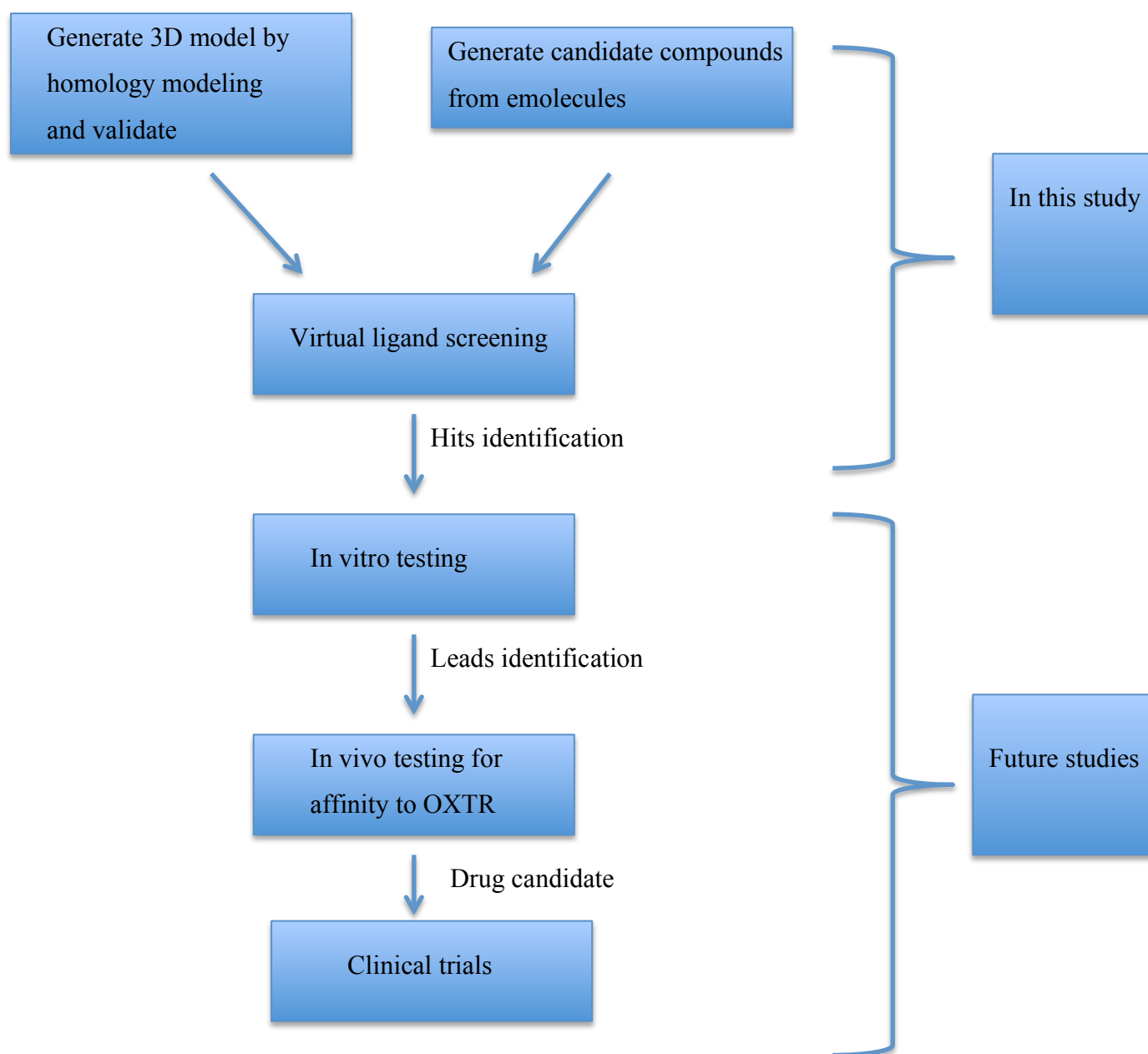


Figure 19 Illustration of future studies

7 Reference

1. Karpenko IA, Margathe J-F, Rodriguez T, Pflimlin E, Dupuis E, Hibert M, Durroux T, Bonnet D. Selective Nonpeptidic Fluorescent Ligands for Oxytocin Receptor: Design, Synthesis, and Application to Time-Resolved FRET Binding Assay. *J Med Chem.* 2015;58(5):2547-52.
2. Feifel D, Shilling PD, Belcher AM. The effects of oxytocin and its analog, carbetocin, on genetic deficits in sensorimotor gating. *Eur Neuropsychopharmacol.* 2012;22(5):374-8.
3. Gimpl G, Fahrenholz F. The Oxytocin Receptor System: Structure, Function, and Regulation. *Physiol Rev.* 2001;81:629-83.
4. Jojart B, Balogh B, Marki A. Modeling the human oxytocin receptor for drug discovery efforts. *Expert Opin Drug Discov.* 2007;2(12):1579-90.
5. Carson DS, Guastella AJ, Taylor ER, McGregor IS. A brief history of oxytocin and its role in modulating psychostimulant effects. *J Psychopharmacol.* 2013;27(3):231-47.
6. Zingg HH, Laporte SA. The oxytocin receptor. *Trends Endocrinol Metab.* 2003;14(5):222-7.
7. Yang Y, Li H, Ward R, Gao L, Wei JF, Xu TR. Novel oxytocin receptor agonists and antagonists: a patent review (2002 - 2013). *Expert Opin Ther Pat.* 2014;24(1):29-46.
8. Barberis C, Mouillac B, Durroux T. Structural bases of vasopressin/oxytocin receptor function. *J Endocrinol.* 1998;156(2):223-9.
9. Koehbach J, Stockner T, Bergmayr C, Muttenthaler M, Gruber CW. Insights into the molecular evolution of oxytocin receptor ligand binding. *Biochem Soc Trans.* 2013;41(1):197-204.
10. Wesley VJ, Hawtin SR, Howard HC, Wheatley M. Agonist-specific, high-affinity binding epitopes are contributed by an arginine in the N-terminus of the human oxytocin receptor. *Biochemistry.* 2002;41(16):5086-92.
11. Hawtin SR, Howard HC, Wheatley M. Identification of an extracellular segment of the oxytocin receptor providing agonist-specific binding epitopes. *Biochem J.* 2001;354(Pt 2):465-72.
12. Chini B, Mouillac B, Balestre MN, Trumpp-Kallmeyer S, Hoflack J, Hibert M, Andriolo M, Pupier S, Jard S, Barberis C. Two aromatic residues regulate the response of the human oxytocin receptor to the partial agonist arginine vasopressin. *FEBS Lett.* 1996;397(2-3):201-6.
13. Gimpl G, Reitz J, Brauer S, Trossen C. Oxytocin receptors: ligand binding, signalling and cholesterol dependence. *Prog Brain Res.* 2008;170:193-204.
14. Gimpl G, Burger K, Politowska E, Ciarkowski J, Fahrenholz F. Oxytocin receptors and cholesterol: interaction and regulation. *Exp Physiol.* 2000;85 41s-9s.

15. Chandonia JM, Brenner SE. Implications of structural genomics target selection strategies: Pfam5000, whole genome, and random approaches. *Proteins*. 2005;58(1):166-79.
16. Krieger E, Nabuurs SB, Vriend G. Homology modeling. *Methods Biochem Anal*. 2003;44:509-23.
17. Michino M, Abola E, Participants GA, Brooks CL, Dixon JS, Moulton J, Stevens RC. Community-wide assessment of GPCR structure modeling and docking understanding. *Nat Rev Drug Discovery*. 2009;8(6):455-63.
18. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004;3(11):935-49.
19. Hart TN, Read RJ. A multiple-start Monte Carlo docking method. *Proteins*. 1992;13(3):206-22.
20. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian J Intern Med*. 2013;4(2):627-35.
21. Klebe G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discov Today*. 2006;11(13-14):580-94.
22. Ballesteros JA, Shi L, Javitch JA. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. *Mol Pharmacol*. 2001;60(1):1-19.
23. Warne T, Moukhametzianov R, Baker JG, Nehme R, Edwards PC, Leslie AG, Schertler GF, Tate CG. The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. *Nature*. 2011;469(7329):241-4.
24. Miller-Gallacher JL, Nehme R, Warne T, Edwards PC, Schertler GF, Leslie AG, Tate CG. The 2.1 Å resolution structure of cyanopindolol-bound beta1-adrenoceptor identifies an intramembrane Na⁺ ion that stabilises the ligand-free receptor. *PLoS One*. 2014;9(3):e92727.
25. Ring AM, Manglik A, Kruse AC, Enos MD, Weis WI, Garcia KC, Kobilka BK. Adrenaline-activated structure of beta2-adrenoceptor stabilized by an engineered nanobody. *Nature*. 2013;502(7472):575-9.
26. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol*. 1994;235(3):983-1002.
27. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283-91.
28. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. *Nature*. 1996;381(6580):272.
29. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993;2(9):1511-9.
30. Vriend G. WHAT IF: a molecular modeling and drug design program. *J Mol Graph*. 1990;8(1):52-6, 29.

31. Rosse G. Pyrazolsulfonamide Agonists of Oxytocin Receptor. *ACS Med Chem Lett.* 2014;5(11):1188-9.
32. Bhattacharya A, Wunderlich Z, Monleon D, Tejero R, Montelione GT. Assessing model accuracy using the homology modeling automatically software. *Proteins: Struct, Funct, Bioinf.* 2008;70(1):105-18.
33. Worth CL, Kleinau G, Krause G. Comparative Sequence and Structural Analyses of G-Protein-Coupled Receptor Crystal Structures and Implications for Molecular Models. *PLoS One.* 2009;4(9):e7011.
34. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. *Proteins.* 1992;12(4):345-64.
35. Shahlaei M, Madadkar-Sobhani A, Mahnam K, Fassihi A, Saghaie L, Mansourian M. Homology modeling of human CCR5 and analysis of its binding properties through molecular docking and molecular dynamics simulation. *Biochim Biophys Acta.* 2011;1808(3):802-17.
36. Slusarz MJ, Slusarz R, Ciarkowski J. Molecular dynamics simulation of human neurohypophyseal hormone receptors complexed with oxytocin-modeling of an activated state. *J Pept Sci.* 2006;12(3):171-9.
37. Schwartz TW, Frimurer TM, Holst B, Rosenkilde MM, Elling CE. Molecular mechanism of 7TM receptor activation--a global toggle switch model. *Annu Rev Pharmacol Toxicol.* 2006;46:481-519.
38. Fanelli F, Barbier P, Zanchetta D, de Benedetti PG, Chini B. Activation mechanism of human oxytocin receptor: a combined study of experimental and computer-simulated mutagenesis. *Mol Pharmacol.* 1999;56(1):214-25.
39. Goodwin JT, Clark DE. In silico predictions of blood-brain barrier penetration: considerations to "keep in mind". *J Pharmacol Exp Ther.* 2005;315(2):477-83.
40. Lobell M, Molnar L, Keseru GM. Recent advances in the prediction of blood-brain partitioning from molecular structure. *J Pharm Sci.* 2003;92(2):360-70.
41. Frantz MC, Rodrigo J, Boudier L, Durroux T, Mouillac B, Hibert M. Subtlety of the structure-affinity and structure-efficacy relationships around a nonpeptide oxytocin receptor agonist. *J Med Chem.* 2010;53(4):1546-62.

