



HAL
open science

Prediction of molecular interactions and physicochemical properties relevant for vasopressin V2 receptor antagonism

Ania de la Nuez Veulens, Yoanna Ginarte, Rolando Fernandez, Fabrice Leclerc, Luis Cabrera

► **To cite this version:**

Ania de la Nuez Veulens, Yoanna Ginarte, Rolando Fernandez, Fabrice Leclerc, Luis Cabrera. Prediction of molecular interactions and physicochemical properties relevant for vasopressin V2 receptor antagonism. *Journal of Molecular Modeling*, 2022, 28 (2), pp.31. 10.1007/s00894-021-05022-6 . hal-03767713

HAL Id: hal-03767713

<https://hal.science/hal-03767713>

Submitted on 23 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Prediction of molecular interactions and physicochemical properties relevant for vasopressin V2 receptor antagonism

Ania de la Nuez Veulens¹, Yoanna M Álvarez Ginarte², Rolando E Rodríguez Fernández³, Fabrice Leclerc⁴, Luis A Montero Cabrera²

1 Department of Biochemistry, Faculty of Biology, University of Havana, Cuba.

2 Laboratory of Computational and Theoretical Chemistry, Faculty of Chemistry, University of Havana, Cuba.

3 DNAdigest Inc., Cambridge UK.

4 I2BC CNRS-CEA-Université Paris Sud, Paris Saclay, France

Corresponding author: Luis A. Montero Cabrera

Telephone:

email: lmc@fq.uh.cu

Abbreviations and Acronyms:

GPCR: G Protein Coupled Receptor; PKD: Polycystic Kidney Disease ;ADPKD: Autosomal Dominant Polycystic Kidney Disease; AVP: arginine vasopressin; V2R vasopressin V2 receptor; SIADH: Syndrome of inappropriate antidiuretic hormone secretion; ADH: antidiuretic hormone; QSAR: Quantitative Structure Activity Relationship; GA: Genetic Algorithm; MLR: Multiple Linear Regression; POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; RMSD: Root mean-square deviation; NPT: constant number of particles, pressure, and temperature; PDB: Protein Data Bank.

ABSTRACT

We have developed two computational approaches, ligand and receptor-based to study the physicochemical properties relevant to the biological activity of vasopressin V2 receptor (V2R) antagonist and eventually to predict the binding mode to V2R of its antagonist. The Quantitative Structure Activity Relationship (QSAR) model so obtained, showed a correlation of the antagonist activity with the hydration energy (E_{H_2O}) as hydrophilic descriptor, and the polarizability (P) as electronic descriptors, appeared to give a positive contribution to the antagonist activity, whereas using the calculated partial charge on atom N7 (q_6) of the common substructure as electronic descriptor yielded a negative contribution to the antagonist activity. To study the molecular interactions, V2R was modeled and further relaxed on a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane using Molecular Dynamics Simulations. Receptor antagonist complexes were obtained by Molecular Docking, and the stability of the selected complexes were studied also using Molecular Dynamics. Amino acid residues Q96, W99, F105, K116, F178, A194, F307 and M311 were identified with the most relevant antagonist-receptor interactions on the studied complexes. The proposed QSAR model could explain the molecular properties relevant to the antagonist activity, contributing to the antagonist-receptor interaction also in agreement with the binding mode of the complexes obtained by Molecular Docking and Molecular Dynamics. These models will be used in further studies looking for potential new V2R antagonist molecules.

Keywords: V2R, AVP, Vasopressin Antagonist, GPCR, QSAR, Docking, Molecular Dynamics Simulations.

1. INTRODUCTION

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is a genetic condition with an incidence of 1:400, to 1:1000 in the world population. Patients develop multiple fluid-filled cyst in both kidneys, increasing the total kidney volume and leading to Chronic Kidney Disease. In contrast to the normal renal cells, the PKD cystic cells have an increased cAMP induced proliferation by activating the Ras/B-Raf/MEK/ERK pathway, mediated by the decrease in intracellular calcium levels [1]. The absence of renal cysts in PCK rats lacking arginine vasopressin (AVP) might indicate that the receptors on the collecting ducts activating other Adenylate Cyclases, do not have a significant role on the generation of cysts [2, 3].

AVP is a nonapeptide hormone consisting of a six amino acid ring closed by a disulphide bridge between cysteines 1 and 6, followed by a tripeptide tail. This hormone is synthesized in the hypothalamus, principally produced by neurons with the cell body within the supraoptic and the paraventricular nuclei, and their axon terminations in the neural lobe of the posterior pituitary gland in which AVP is released into the circulation [4]. The primary function of AVP is to maintain body fluid balance by keeping plasma osmolality within narrow limits [5]. Increase in plasma osmolality or decrease in plasma volume trigger its release to induce expression of water transport proteins in the late distal tubule and collecting ducts of the kidneys, to increase water reabsorption [5]. Due to its role in the regulation of osmolarity by increasing the ability of the kidney to reabsorb water reducing the urinary volume, it is also known as Antidiuretic Hormone (ADH).

AVP physiological roles are mediated by three receptor subtypes V1a, V1b (also called V3), and V2 all belong to vasopressin/oxytocin receptor family and they are class-A G-protein coupled receptor (GPCR). The V1a receptors are mainly distributed on vascular smooth muscle, but also present in myocardium, platelets, and hepatocytes. V1a stimulation is associated with vasoconstriction and cardiac hypertrophy, together with platelet aggregation, and glycogenolysis [4, 6, 7]. The V1b receptors have little selective distribution and their activation is part of the adaptive reaction to stress, leading to stimulation of adrenocorticotrophic hormone and endorphin release [4, 6]. The activity of each receptor is mediated by G proteins which activate a phosphatidyl-inositol-calcium second messenger system.

The V2 receptor (V2R) is expressed predominantly in the principal cells of the renal collecting duct system, in which its activation leads to increased resorption of free water [4–6]. V2R is the major activator of adenylyl cyclase signaling pathway in principals cell of collecting ducts in kidney. The increase of cAMP intracellular concentration by the activation of V2R promote proliferation in PKD cystic cells, suggesting that V2R antagonists can be used as treatment for PKD to retard development and growth of the cysts [8–10].

Selective peptide antagonist of V2R were developed [11, 12] but these efforts have encountered many obstacles due to the residual agonist activity, heterogeneity in species response and very low oral bioavailability limiting their clinical use [13]. These limitations make the development of new non-peptide V2R antagonists more attractive.

Orally and intravenously active non-peptide vasopressin receptor antagonists are called vaptans. The first success in this field was mozavaptan (OPC-31260), a benzazepine derivative and a potent, selective, competitive and orally active vasopressin V2 receptor antagonist [14], soon followed by its use in humans [15] and the first to gain approval for clinical use in Japan since 2006 for the treatment of tumor-associated to Syndrome of inappropriate Antidiuretic Hormone secretion (SIADH).

Among the non-peptide V2R antagonists developed and experimentally tested, only two compounds of this class have been approved in the United States, Canada and the European Union [13]. The U.S. Food and Drug Administration (FDA) approved conivaptan and tolvaptan for euvoletic and hypervolemic hyponatremia [16]. Tolvaptan is also approved to slow kidney function decline in adults at risk of rapidly progressing Autosomal Dominant Polycystic Kidney Disease (ADPKD), the only drug approved to treat this condition so far [16].

Besides the retardation of progressive renal failure in ADPKD and the treatment for euvoletic or hypervolemic hyponatremia, experiments show that V2R antagonists can be used for rescue treatment in Congenital Nephrogenic Diabetes Insipidus [17], treatment of diabetic nephropathy [18], congestive heart failure [19], and also in the prevention of ascites formation in cirrhosis [20]. Other indications for treatment with vasopressin-receptor antagonists will probably emerge.

In this work we have used a computational modeling approach to study a family of non-peptide V2R antagonists, with their IC_{50} being determined experimentally, identifying physicochemical properties relevant to the biological activity of these compounds and relevant interactions with V2R.

2. METHODS

2.1. Data set of V2R Antagonist

To obtain a reliable QSAR model we used chemical information from the assays AID-217680, AID-2176881 and AID-483985, all in the PubChem's Bio Assay Database (<https://pubchem.ncbi.nlm.nih.gov>) completing a series of 53 antagonists. In these assays, the biological activity at V2R was assessed as the displacement of [3H]-AVP from its AVP-V2R binding site and the inhibition of intracellular cAMP accumulation. The IC_{50} value is the concentration of compound which inhibits [3H]-AVP binding by 50%. In our study, the negative

logarithm of the biological activity, pIC_{50} , was used as the dependent variable to determine QSAR correlation equations. The 3D structure of each antagonist were generated from its SMILES in Pubchem database.

2.2. Estimation of molecular properties

The specific action of drugs depends on many intrinsic features such as hydrophobic, electronic, and steric properties. In a QSAR model, the biological activity is expressed as a function of molecular descriptors. A molecular descriptor encodes as a number, the result of a mathematical and logical procedure using the information of specific properties of molecules. In this study we calculated as a hydrophobic descriptor, the logarithm of the octanol/water partition coefficient (LogP) and hydration energy; as steric descriptors: approximate surface area (ASA), grid surface area (GSA), molar volume (MV)[21, 22], and molar refractivity (MR); as for electronic descriptors: polarizability (P) [23], dipole moment (μ), total energy (TE), highest occupied molecular orbital eigenvalue (eHOMO), lowest unoccupied molecular orbital eigenvalue (eLUMO), partial atomic charges of the pharmacophore atoms (q1 to q11), electrophilicity index (ω), chemical hardness (η), chemical softness (s) [24]. Electronic descriptors were calculated by Kohn-Sham's DFT *B3LYP/6-31G* method as included in Gaussian 09 program routines [25]. The other descriptors were calculated with QSAR properties available in Hyperchem v8 software [26].

2.3. Cluster analysis

Cluster analysis is used in QSAR models to build the training and test sets as well as to determine the structural diversity of the dataset. In cluster analysis, the antagonists were classified in groups, called clusters, with a relative homogeneity. The structural diversity or similarity between the compounds is determined by calculating the Euclidean distance between each couple of objects: the smaller the distance, the more of the objects are considered similar to each other [27]. To check the structural diversity of the dataset and to define the number of possibles clusters, a hierarchical cluster analysis of these molecules was performed using k-NNCA algorithm to construct the dendrogram. The complete linkage distance (Euclidean metric) was used as the connection function to merge the objects into clusters. The complete linkage measures the proximity between two groups, calculating the distance among the farthest objects, or the similarity among the objects with lesser similarities. The Euclidean distance is the square root of the sum of the squared differences among the values of two objects for each variable [28, 29].

To select the training and test sets we used the k-mean cluster algorithm (k-MCA). Such algorithms use a switching method to divide N data points into k groups (clusters) to minimize the sum of

distances/dissimilarities among the objects within the same cluster. The k-mean approach requires that k (the number of clusters) must be known before clustering [29]. The k values were set taking into account the dendrogram obtained for the first cluster analysis. Both hierarchical and partitional (non-hierarchical) cluster analyses were implemented using the STATISTICA 8 software [30]. After the cluster analysis, the compounds were separated in two sets: 80% of compounds in each cluster were selected for the training set and 20% of each cluster for the test set. The training set was used to develop the QSAR model and the test set was used for external cross-validation of the model.

2.4. QSAR model

A correlation matrix was performed to determine among the calculated molecular descriptors the ones that do not correlate to each other. A Genetic Algorithm (GA) was used as a metaheuristic method for the molecular descriptors selection and optimization of the functions [31–34]. The length of the equation was set for three terms and a constant, and the GA was used for input selection to establish which of the descriptors will have the best multiple linear regression (MLR). Several statistical parameters were employed to validate the model. A good QSAR model should have the highest squared correlation coefficient, R^2 and Fisher-test, with the lowest standard deviation (S). The P-value is another important parameter used for modeling validation and it should be lower than 0.01. The predictive power of the model was then determined by examining the leave-one-out (LOO) cross-validation (q^2). The q^2 is known as the predictive variance, with a value higher than 0.5. To validate the QSAR model, an external prediction test set of compounds (in the model range) was used, as the predictive ability of a QSAR model shall only be estimated using an external test set [35, 36]. All the procedures used to build the QSAR model were performed with BuildQSAR software [37] and validated with STATISTICA 8 software [30].

2.5. The applicability domain

The applicability domain is the theoretical region of the chemical space, defined by the model descriptors and the modeled response, therefore by the nature of the compounds in the training and test sets, as represented in each model by specific molecular descriptors. The applicability domain of a QSAR model is “the range within which it tolerates a new molecule“ [28], a QSAR model is only valid within the same domain for which it was developed. Even if the models are developed for the same chemical structures, the applicability domain for new structures can differ from model to model, depending on specific descriptors.

In multiple predictor models, performing simple single-variable range checks is not sufficient to verify the applicability domain. For MLR, one of the most used approaches with normally

distributed data for a multiple predictor problem is a distance-based measure like the leverage (h). As the leverage of a compound measures its influence on the model, it becomes possible to verify whether a new chemical will fit within the structural model domain. The leverage used as a quantitative measure of model applicability domain is also suitable for evaluating the degree of extrapolation, which represents a sort of compound distance from the model experimental space. The warning leverage (h^*) is a critical value or cut-off. Predictions should be considered unreliable for compounds with high leverage ($h > h^*$, being the critical value $h^* = 3p'/n$, where p' is the number of model variables plus one, and n is the number of the compounds) [28].

2.6. V2R modeling

GPCRdb template tool (https://gpcrdb.org/structure/template_selection) was used to identify the possible templates for V2R and the human OX2 orexin receptor (PDB ID 4S0V) was selected for V2R modeling. The alignment between the template sequence and V2R was made with a Clustal X v2.1 profile [38] taking into account the alignment already obtained with the GPCRdb template tool. The model was built by homology with YASARA program v12.8.26 [39]. YASARA uses knowledge-based energies to validate the receptor model normalizing them to remove the dependencies on the size and shape of the protein, and also on its amino acid composition, obtaining estimates for the expected average energy and its standard deviation from gold-standard reference structures. Then it calculates how many standard deviations it is away from the average, thereby obtaining a Z-score to evaluate the quality of the models.

To minimize and equilibrate the receptor model, it was inserted in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) membrane patch, generated with the VMD [40] Build Membrane plugin, which mimics its natural environment. POPC was selected because its abundance in biological membranes and because it does not introduce any curvature in the structure. The receptor model was oriented on the membrane, according to the orientation of the orexin receptor (ID PDB 4S0V) on the membrane in the MemProtMD database [41]. The membrane with the receptor inserted was oriented in the XY plane. The system was solvated in a $93 \times 92 \times 113$ Å periodic box of TIP3P water and NaCl was added at physiological concentration, neutralizing the system. The energy of the system was minimized with 1000 steps of conjugate gradient and a further equilibration for 10ns (0.5ns to 300K lipids tails; 0.5ns to 300K membrane; 9ns to 310K, on the whole system) was performed with NAMD2 [42] with an integration time step of 2 fs. A Langevin thermostat and barostat were used to maintain an NPT system, the cut-off for non-bonding interactions was 12 Å, a smooth switching function at 10 Å was used for van der Waals interactions, and Particle Mesh Ewald (PME) for electrostatic interactions. The membrane parameters were checked with MEMBPLUGIN [43] of VMD.

2.7. Molecular docking

The receptor model obtained in the previous simulation was used to perform docking studies to predict the binding modes of three antagonist of the studied family. The receptor model and the antagonists were prepared using AutoDockTools [44] to perform the molecular docking with AutoDock Vina [45], where the antagonist and the sidechains of the receptor residues (Q96, W99, F105, K116, and F307) were flexible. The search space was restricted to a 28 x 20 x 14 Å box. The default parameters for configuration files were used in Autodock Vina, running it 5 times and saving 10 conformations of each compound for each run to generate a total of 50 conformations for each compound. The docking results were visually analyzed using UCSF Chimera [46]. All docking results were clustered using a tolerance value of 2.0 Å RMSD and three representative orientations in the binding site were identified to select one conformation per compound. The interactions between the selected antagonist conformation and the V2R were analyzed using BINANA [47].

2.8. Molecular dynamic simulations of complexes in POPC

The topologies and parameters of the selected antagonists for CHARMM 36 force field [48] were generated using CHARMM-GUI [49, 50]. The complex in the POPC membrane was minimized (10000 steps) and equilibrated for 100 ps at 310K with NAMD2. A molecular dynamic simulation was performed for 50 ns (310K, NTP and constant area) with NAMD2, saving frames and calculating energy every 5000 steps.

For the equilibration and the production simulations, a Langevin thermostat and barostat were used to maintain an NPT system, the cut-off for non-bonding interactions was 12 Å, a smooth switching function at 10 Å was used for van der Waals interactions and Particle Mesh Ewald (PME) for electrostatics interactions. The integration time step was 2 fs. VMD [40] was used for the analysis and visualization of the molecular dynamic simulations.

2.9. Complex free energy calculations using linear interaction energy methods

The linear interaction energy (LIE) method was used to estimate the free energy of antagonist-receptor binding. For this purpose, in addition to the previous simulation of the complex in the membrane in a cubic water box, a second simulation of the antagonist only in the water box is needed, which was carried out using the same parameters as the simulation of the antagonist-receptor complexes. Eq.1 shows the improved LIE formula suggested by Almlöf et al. [51, 52] and takes into account the intra-ligand electrostatic interactions.

$$\Delta G_{bind} = \beta (\Delta \langle V^{el}_{L-S} \rangle + \Delta \langle V^{el}_{L-L} \rangle) + \alpha \Delta \langle V^{vdw}_{L-S} \rangle + \gamma \quad (1)$$

Where $\langle V^{el}_{L-S} \rangle$ and $\langle V^{vdw}_{L-S} \rangle$ are MD-generated interaction energy averages from the non-bonded electrostatic and van der Waals interactions of the ligand with its surrounding environment (s). $\langle V^{el}_{L-L} \rangle$ is the electrostatic intramolecular ligand-ligand average energy. The Δ 's denote the change in average values when transferring the ligand from solution (free state) into the binding site of the solvated receptor (bound state). Coefficients α and β are scaling factors for the energy terms, while γ is an empirical constant. In this study, α was considered as 0.18 which is considered to be a robust value from previous works [51–53]. The β specific values for each antagonist was calculated using the parameterization model E proposed by Almlöf *et al.* [51, 52] (Eq.2).

$$\beta_{FEP} = \beta_0 + \frac{\sum_i w_i \Delta \beta_i}{\sum_i w_i} \quad (2)$$

Where w_i , β_0 , and $\Delta \beta_i$, were calculated from explicit solvent FEP calculations of single chemical group ($w_i = 1$ if group is neutral and 11 if it is an anion or a cation), $\beta_0 = 0.43$ and $\Delta \beta_i$ was obtained by the model proposed by Almlöf *et al* [51].

The balance (difference) between the electrostatic (polar) and the van der Waals (nonpolar) contributions to the free energy binding in the LIE method was defined as the parameter D (Eq. 3)

$$D = \beta (\Delta \langle V^{el}_{L-S} \rangle + \Delta \langle V^{el}_{L-L} \rangle) - \alpha \Delta \langle V^{vdw}_{L-S} \rangle \quad [\text{kcal/mol}] \quad (3)$$

LIE-D is an approach based on the linear correlation between the γ coefficient and the D parameter that accounts for the balance (difference) between the polar and nonpolar binding free energy contribution. The relationship between the γ coefficient and D parameter takes the form:

$$\gamma = f x D + g \quad [\text{kcal/mol}] \quad (4)$$

The values of f and g were estimated by Miranda *et al.* [52] as -0.95 and -2.06 respectively.

3. RESULTS AND DISCUSSION

3.1. Construction of training and test sets using Cluster Analysis

We selected a series of 53 compound antagonists of V2R to construct the training and test sets. All the selected molecules have the same core substructure Figure 1, but show structure variability due to substituent structural diversity. All compounds have a common substructure (4-formamido-benzamide) remarked in a rectangle in the top panel of Figure 1. The nitrogen of the benzamide in the common substructure (represented the partial charge q1) is part of a benzazepine, benzene-piperidine, or benzoxazine condensed ring. The R1 substituent is generally a ring except for

compounds A03 and A07. We display in Table 1 the compounds we used in this study with PubChem ID and the experimental biological activities of V2R antagonists' activity expressed as IC₅₀ and pIC₅₀.

To classify the molecules of the datasets, depending on their structural variability we performed a hierarchical cluster analysis, the resulting dendrogram was constructed using the Euclidean distance (x-axis) and the complete linkage (y-axis), illustrating the results of the k-NNCA developed in this dataset. The dendrogram shows 6 different subsets demonstrating the molecular variability among the compounds of this dataset (Figure 2). To evaluate the output dendrogram and to split the whole dataset into training and test sets, we performed a k-mean cluster analysis (k-MCA) [54].

The selection of the training and test sets was carried out by randomly taking molecules belonging to each cluster. From the initial 53 compounds, 42 (80% of the dataset) were chosen to form the training set and the remaining 11 compounds, (20% of the dataset) were used as a test set for the external cross-validation of the model.

3.2. Development and validation of the QSAR model

GA combined with MLR is widely used for QSAR and QSPR studies [31–34]. In this method, a GA is performed to search the feature space and select the major descriptors relevant to the activities or properties of the compounds. This method can deal efficiently with a large search space, and it has fewer chances to only find a local optimal solution than other algorithms. GA is a well-estimated method for parameter selection and to overcome the shortages of MLR in variable selection. After a GA, the MLR is employed to correlate the selected descriptors with the activity values using a classic regression method to yield the explicit equations.

The variables selected by the genetic algorithm as the best model of V2R antagonist activity are shown in equation 5. To further validate the variables thus obtained, we performed an MLR analysis of the 43 compounds on the initial training set, with the 11 compound test set for the external cross-validation.

$$pIC_{50} = -7.968 (\pm 3.584) \mathbf{q6} + 0.095 (\pm 0.059) \mathbf{E_{H2O}} + 0.161 (\pm 0.027) \mathbf{P} - 5.842 (\pm 2.894) \quad (5)$$

$$n = 43; R = 0.89; R^2 = 0.80; s = 0.40; F = 53,61; p < 0.0001; q^2 = 0.75$$

Test set:

$$n = 11; R = 0.86; R^2 = 0.74; s = 0.41; F = 25.04; p < 0.0007; q^2 = 0.56$$

The R^2 (R-square statistic or coefficient of determination) indicates that the model could explain 80 % of the variance for the experimental values of pIC_{50} . The model shows a q^2 of 0.75. This value of more than 0.5 could be considered as proof of the high predictive ability of the model, along with the good prediction of the test set ($R^2 = 0.74$). The good R^2 and q^2 values obtained in equation 5 for both training and test set can be explained with the experimental values for all the compounds of the series. The calculated values for pIC_{50} , are highly similar to the experimental, sustaining the reliability of the QSAR model (Figure 3, Table 2).

In the correlation study with the calculated descriptors, a low correlation was observed between the variables, indicating the reliable information content on each term in the equation (Table 4). The selected variables by the genetic algorithm were P (polarizability) E_{H_2O} (hydration energy), and q_6 (partial charge of nitrogen in the common substructure 4-formamidobenzamide (Figure 1)). For each one of the variables, the coefficients were significant (Table 5), indicating their relative contribution to the combined prediction of the biological activity as the dependent variable. Calculating the value of the coefficients on the regression analysis we ensure a good prediction starting from the group of the independent variables (q_6 , E_{H_2O} and P), facilitating the interpretation of the independent influence of each variable on the final equation.

The variable q_6 represents the partial charge of the N7 in the 4-formamidobenzamide common substructure, involved on an amide bond associated with the variable zone of the compounds. The partial charge q_6 is the most negative, with the highest module value for all the studied charges, having a fully negative value range and a negative coefficient on equation 5, could indicate the favorable tendency of an increased antagonist activity with more negative values of q_6 . It could be explained by the fact that N7 is involved in a hydrogen bond, or because it just reflects the variation of partial charge depending on the nature of substituent R1. The partial charge q_6 , calculated only for the common substructure is -0.712 and the substituent on R1 is making it more negative, except for compounds A53, A24, and A12. In the case of A12, the N7 is substituted by a methyl, which has a strong inductive effect (+i) over the nitrogen. Compounds A53 and A24 both have a benzene ring with a nitro substituting in the ortho position as R1. The common substructure in most compounds is formed entirely by a conjugated system with a substituted benzene ring as R1, which might contribute to a whole conjugated system. If we compare the compounds by the position of the substituent on the benzene ring at R1, we observe a lower partial charge on q_6 associated with an ortho substituent, calling for a combination of steric and electronic factors, where the ortho substituent could break the conjugation planarity.

Hydration energy (E_{H_2O}) is the amount of energy released when one mole of a compound is hydrated, and represents the measure of the water molecules affinity for the compound. More negative hydration energy values could be associated with more polar groups in the compound, and

less negative hydration energy could be attributed to the presence of a higher number of nonpolar groups [55]. The range of values for this variable in the data set is negative and it has a positive coefficient on equation 5, thus indicating that more negative values of hydration energy are unfavorable for the antagonist activity, suggesting a binding site with possible hydrophobic interactions, as more hydrophilic compounds are shown unfavorable for the activity.

Polarizability refers to the tendency of any compound to acquire an electric dipole moment in proportion to an applied electric field, on our model the Polarizability component is having a correlation coefficient of 0.7 with the pIC_{50} , making it the descriptor with the highest correlation to the activity, being the other variables the fine adjustments necessary to improve the model in general. The range of polarizability values in the data set is positive and has a positive coefficient in equation 5, so it has a favorable contribution to antagonist activity.

As mentioned above, the compounds in the studied dataset have conjugated systems in their structure, and systems with delocalized π electrons exhibit high polarizabilities. The aromatic systems' planarity with their high polarizability and multipole moment, are all factors of key importance for the 3D architecture of aromatic complexes [56]. Soft interactions like dispersion, are predominant in stacking and can be estimated from the polarizability [57]. In general, the studied compounds have three aromatic rings in their structure that could be involved in pi-interactions with the residues in the receptor-binding site.

In general, the obtained QSAR model provides indications that the binding mode of V2R antagonists might fundamentally be involving hydrophobic and electron density interactions.

3.3. The applicability domain (AD) of the QSAR model

A QSAR model needs to show not only a good accuracy, but also some reliability for predictions of new compounds. These models cannot be universal and should be constrained to a defined chemical space, commonly known as the applicability domain (AD). The AD can be described as the physicochemical, structural or biological spatial information based on which the model training set is developed. The QSAR model applies to make predictions for new compounds within the specific domain [57]; in summary, the AD is the degree to which a QSAR model tolerates (reliably) new compounds.

A crucial problem in chemometrics and QSAR studies is the definition of the AD with a regression model. We will define it here as a squared area within ± 2 bands for standardized residuals and a leverage threshold of $h=0.23$ for inhibitory activity (Eq. 5). Thus, compounds with standardized residuals greater than 2 standard deviations will be considered unreliable. For the graphical visualization of outliers for the response (standardized residuals > 2) or for the structure (leverage $>$

0.23) in the regression model, the Williams plot for Eq. 5 is shown in Figure 4. Of the 53 compounds in the dataset, only two compounds (A02 and A12) have a leverage higher than the critical value.

A02 (conivaptan) has the highest value of polarizability (57.51) of the dataset, while the other compounds are between 36.87 to 55.95. A02 shows a diphenyl moiety as substituent of the amide in the common substructure, while the other compounds exhibit only a single aromatic ring or an aliphatic substituent, A02 also have a condensed 3 ring system of 3,4,5,6-tetrahydroimidazol[4,5-d][1]benzazepine, while the other compounds have only a 2 ring system. The presence of extra rings on A02, might account for the increase on the polarizability for this compound with a different electronic structure than the rest.

A12 shows the highest value of q_6 (-0.515) of the dataset, while the other compounds are between -0.705 to -0.844, and it also has the highest value of hydration energy (E_{H_2O}), with -3.44 kcal/mol, while the other compounds are between -5.05 to -14.07 kcal/mol. Compound A12 show a minor difference in the common substructure with all the antagonists of this family having a methyl group as substituent for the N amide, directly altering the partial charge (q_6) of N7 on the 4-formamidobenzamide and increasing the hydration energy value. The structure of A12 and A01 differ only in the aforementioned methyl group, A12 has a difference of 0.203 on q_6 , and of 2.36 kcal/mol in the value of hydration energy compared to A01, evidencing the influence that this single methyl group can have.

3.4. V2R modeling

We selected the human OX2 orexin receptor (PDB ID 4S0V) as the template for V2R modeling, as suggested by the GPCRdb template tool. The selected template has 27% of identity and 46% of similarity with V2R. In the corresponding alignment, the fragments corresponding to transmembrane helices and the conserved motifs are preserved (Figure 5). In both the OX2 receptor and V2R, the natural ligand is a peptide, and the selected structure has an antagonist bound being on an inactive conformation, suitable to study the binding modes of antagonists to the V2 receptor.

To relax the obtained model in a more natural environment, it was minimized and equilibrated for 10ns in a POPC membrane, solvated and with NaCl added at physiological concentration (0.15M). At the end of the simulation the membrane parameters, like thicknesses and per lipid area, were calculated to check the correct packing. The membrane thickness (distances between phosphates of each monolayer) was 38.91 Å, and the per lipid area was 64.54 Å². These parameters are reasonable for a POPC membrane at 310K, according to experimental parameters obtained at different temperatures [59].

During the current work, another X-ray structure from the class A of GPCR was released in the Protein Data Bank (PDB ID 6TPK): the oxytocin receptor (OXTR) which is also a member of the vasopressin receptor family. Although OXTR exhibits slightly better sequence identity and similarity: 41% and 56% respectively, its lower resolution (3.20Å versus 2.50Å), a missing region (loop and helix 8), and a shorter loop (...) make it a less relevant template. However, in an effort for further validation, the V2R model was compared to OXTR; the superimposition between the V2R model and OXTR is shown in Figure 6, where we considered the two conformations of V2R before and after membrane relaxation.

The RMSD values, using OXTR as reference and comparing the model before and after the membrane relaxation were 1.02Å and 1.15Å respectively. The main differences between OXTR and the models were that OXTR lacks the ICL3, a long intracellular loop involved in the interaction with the G-protein usually missing in GPCR solved structures and the helix 8, parallel to the membrane and useful for orienting the receptor in the membrane. Comparing the bundled helices, the main difference relative to the binding site is that the TM2 of relaxed model in the membrane is in the same position as that of the OXTR, while the model before relaxing has the TM2 slightly tilted towards the interior of the cavity decreasing its volume. The difference in the orientation of TM2 in the model before relaxation could be caused by the difference in the proline position in this transmembrane section between V2R and the template (Figure 5). Proline in the middle of alpha helices cause a kink, by being unable to complete the H-bonding chain of the helix and because of steric and/or rotameric effects keeping it out from the preferred helical geometry [60]. Proline one position earlier on the sequence of the template TM2 with respect to that of the model, can affect the orientation of the kink and result on a different orientation in the model. This odd orientation could later be corrected during the relaxation in the membrane showing this process as very favorable.

The comparison between the V2R model after relaxation and OXTR brings more confidence in the model's quality and the protocol used for relaxation.

3.5. V2R-antagonist complexes

Visual inspection of the binding site revealed that the side chains of residues W99 and F307 are occluding its entrance, therefore these two residues were considered as flexible for the molecular docking. To improve docking results other residues of the binding site (Q96, F105 and K116) were also considered as flexible.

The antagonists selected as ligands for the molecular docking were: mozavaptan (A01), conivaptan (A02), and tolvaptan. Tolvaptan also shares the common substructure of the studied compound

series for the QSAR model and it is the only drug approved to treat Polycystic Kidney Disease. All the rotatable bonds of the ligands were flexible.

The 50 complexes for each antagonist obtained by molecular docking were clustered for analysis. Three orientations of the antagonists in the binding site were identified and shown in Figure 7. The first with the condensed ring of the antagonist toward TM2 and TM7 (CR-27), the second with the condensed ring towards TM5 and TM6 (CR-56) and the last one with the condensed ring towards the entrance of the cavity (CR-UP). The binding energies of the complexes obtained for the three identified conformations differ by less than 1 kcal/mol in each of the antagonists studied. This difference in the energy value is lower than the standard error of the Autodock Vina scoring function [45], so we might need more studies to select the best conformation.

We expect a good antagonist to bind with high affinity to the receptor binding site but failing to activate it, blocking the access of any agonist to the binding site. A study with meta-dynamics enhanced sampling revealed the existence of three binding sub-sites for V2R, proposed to respond to the vasopressin entry pathway [61]. The compounds that bound in both the vestibule and the intermediate sites block the access to the orthosteric site so that an agonist will never be able to bind, if there is an antagonist already bound to any of the non-activating sites. Two of the antagonists studied by Saleh *et al.* [61], with high structural similarity to those in this study, were predicted to bind to vestibule site for V2R and intermediate site for V1aR, so it is to be expected that the antagonists in our study are located in one of these sites. Therefore, we eliminated from our subsequent analysis the CR-UP conformation, showing the antagonist penetrating deeper into the cavity with part of it located in the orthosteric site.

To study the stability of the compounds in the binding site and make a better estimation of the binding energy, a molecular dynamic simulation of the best complex of the two remaining orientation was performed. Figure 8 shows the RMSD for the two different conformations of the three studied antagonists. The conformations of the studied antagonists tend to stabilize along the molecular dynamics simulations, being CR-27 the conformation with lower RMSD for each of the antagonist. For mozavaptan, the conformation CR-56 is the conformation with more fluctuations along the trajectory and it has the highest RMSD value among all antagonist conformations. The change in the antagonist's conformation for the six complexes in the molecular dynamics simulation is shown in Figure 9. The two representative conformations for each antagonist are represented from left to right (CR-27 and CR-56 respectively). Tolvaptan is represented in green, conivaptan in blue and mozavaptan in pink. The starting conformation at 0 ns is represented by the light colored ligand and the final conformation at 50 ns by the dark-colored ligand. Mozavaptan CR-56 conformation showed a significant change in the orientation of the antagonist with respect to the starting conformation, also reflected in the high RMSD value observed for this conformation. For

this reason, this conformation was eliminated from the subsequent analysis.

In order to predict which binding conformation could be the best for each antagonist, we estimated the binding free energy variation using LIE-D method. This method is flexible enough to consider different interaction patterns, even though the ligands share some common chemical scaffolds and are bound to the same protein receptor [52].

For tolvaptan, the conformation with the best binding free energy is tolvaptan CR-27 (-14.34 kcal/mol) with over 2 kcal/mol of difference with CR-56, that was thus discarded from further analysis. The conformations of conivaptan have similar energies between them, so it is not possible to select which of the two conformations is the best using this criterion. The inhibition constant (K_i) was calculated from the estimated binding free energy for these conformations and compared with experimental values (Table 6). The values of the calculated K_i follow the same trend than the experimental inhibition constants [62, 63], except for the tolvaptan, with the K_i for conformation CR-27 one order lower than the experimental value and the K_i conformation CR-56 one order higher, suggesting as expected that CR-27 was the best conformation for tolvaptan.

The binding free energy and K_i values obtained for the different conformations of the antagonists bound to V2R allowed us to select the CR-27 conformations (or the CR-56 conformation for conivaptan only) as the possible binding modes of these antagonists to V2R. From these conformations, the analysis of the main antagonist-receptor complex interactions can be carried out.

3.6. Interactions analysis of the best antagonist-V2R complexes

The antagonists must block the access or interact with those residues favoring the union of any agonist for the receptor activation, and/or also sterically blocking the residues involved in triggering the activation mechanism. The most relevant contacts between the antagonist and the receptor are summarized in Table 7. The interactions observed for the mozavaptan and tolvaptan complexes are very similar, while conivaptan interacts with a greater number of residues, since it is a compound with a greater volume than those mentioned above.

In the analysis of the obtained complexes, there are some common interaction involving the hydrophobic (C192, A194, L310, M311), aromatic (W99, F105, F178, F307) and polar (Q96, K116, Q119, Q291) residues. The residues Q96, Q119, Q291, K116 are highly conserved in all the AVP and OXT receptors' family, and are known to have a key role in agonist binding [64]. Previous studies with V1aR suggest that the residues Q96, Q119, Q291 are K116 are specifically involved with the ligand binding process but do not intrinsically modulate the efficacy of the functional response [64]. An analysis of the presence of H-bonds along the molecular dynamic simulation was performed to study the nature of the interactions between the antagonists and these polar residues of the receptor. The percentage of frames of the trajectory with a determined number of H-bonds is

shown in the Table 8. In all the studied conformations the occurrence of H-bonds is low, approximately between 13 and 38%, although interactions with some polar residues are observed in greater percentages of the trajectory (Table 7), which suggests other kind of interactions.

The antagonists interact with the aromatic residues W99, F105, F178 and F307. These residues are not directly involved in the receptor activation and they are near or at the entrance of the binding site, suggesting that the antagonist binding site might not be located deep into the cavity. The fact that the antagonists interact with residues near the entrance of the cavity agrees with the binding sites predicted by Saleh *et al.* [61].

Some of these residues have been shown to be involved in vasopressin binding. W99 plays a fundamental role in stabilizing the vasopressin/receptor interactions responsible for the high-affinity binding of agonists to the V2 receptor and receptor selectivity. A mutation of W99 (W99R) greatly impaired the binding properties of the receptor and had a minor effect on its intracellular routing [65]. Other important residue for AVP binding is F105, the mutation F105V was reported to show cell surface expression and a maximal AVP-induced cAMP formation (V_{max}) comparable to the wild type, but with a reduced ligand binding ability [66, 67].

An interesting interaction for the V2R antagonists is with F307, a non-conserved residue in vasopressin/oxytocin family since V1aR has a threonine in this position. The relevance of this interaction is because some antagonists could bind to both V2R and V1aR due to the similarity of its binding site, but the interaction with F307 would be unique for V2R making it attractive for the design of antagonists having less selectivity for V1aR.

Other antagonist-receptor interactions found were with residues C192, A194 and M311. While C192 and A194 are conserved among the entire family, M311 is not seeming to cooperate in the selective binding of some antagonists [68]. The M311V mutation in the TM7 of V2R has impaired the ligand capacity and binding [69] suggesting that in V2R the residue M311 could take part in the binding of peptide agonists [67, 70]

Taking into account the interactions and the estimated binding free energy, we considered CR-27 conformation (and CR-56 for conivaptan) as the best for the three antagonists studied by molecular dynamics. Figure 10 shows the CR-27 conformation of tolvaptan in complex with V2R.

In general, the main interactions observed here are involved in the binding of ligands to V2R, but are not involved on the receptor activation, which suggest that the studied conformations of the antagonists can block the binding of agonists and unable to activate the receptor. The presence of few H-bonds with polar residues and the other interactions observed are with aromatic residues and non-polar residues suggest that the main antagonist-receptor interactions are mostly hydrophobic in nature and could involve pi-clouds.

4. CONCLUSION

In summary, two computational approaches, ligand and receptor based were developed to study the physicochemical properties relevant to the biological activity of V2R antagonists and to predict their binding mode to V2R. The proposed QSAR model allows us to clarify the contribution of three molecular descriptors to the biological activity. Our model described the antagonist activity in correlation with polarizability, hydration energy and partial charge on atom N7, explaining the molecular properties contributing to the antagonist-receptor interaction and relevant to the antagonist activity, which is also in agreement with the binding modes for the complexes obtained by molecular docking and molecular dynamics simulation.

A good quality model based on the structure of OX2 orexin receptor was obtained and used to estimate the antagonist orientations in the binding site of V2R. The conformations of studied antagonist were analyzed by molecular dynamics. In general, the CR-27 conformation is considered as the best conformation for the antagonist binding (through interaction analysis and binding free energy estimation). Most of the relevant interactions observed along the molecular dynamics simulation involve the electronic density by the interaction of the antagonist rings mainly with the aromatic residues (W99, F105, F178 and F307), which is in correspondence with what is expected according to the polarizability variable of our QSAR model. Other relevant interactions are hydrophobic in nature (A194 and M311) which agrees with the expected effect of the hydration energy to the antagonist activity in the QSAR model.

The results obtained by both developed approaches are in fair agreement and contribute to a better understanding of V2R antagonism. These results represent a step forward for the efficient search of potential new V2R antagonist molecules.

Acknowledgment

The authors thank the assistance from the Department of Applied Physical Chemistry of Autonomous University of Madrid, Spain, and the support of the Faculties of Chemistry and Biology from the University of Havana, Cuba. This research was supported through high-performance computing resources provided by the I2BC CNRS-CEA-Université Paris Sud, Paris Saclay, France.

Conflicts of interest

The authors confirm the article content has no conflicts of interest.

5. REFERENCES

1. Yamaguchi T, Wallace DP, Magenheimer BS, et al (2004) Calcium restriction allows cAMP activation of the B-Raf/ERK pathway, switching cells to a cAMP-dependent growth-stimulated phenotype. *J Biol Chem* 279:40419–40430. <https://doi.org/10.1074/jbc.M405079200>
2. Torres VE, Wang X, Qian Q, et al (2004) Effective treatment of an orthologous model of autosomal dominant polycystic kidney disease. *Nat Med* 10:363–364. <https://doi.org/10.1038/nm1004>
3. Grantham JJ (2014) Rationale for early treatment of polycystic kidney disease. *Pediatr Nephrol Berl Ger*. <https://doi.org/10.1007/s00467-014-2882-8>
4. Decaux G, Soupart A, Vassart G (2008) Non-peptide arginine-vasopressin antagonists: the vaptans. *The Lancet* 371:1624–1632. [https://doi.org/10.1016/S0140-6736\(08\)60695-9](https://doi.org/10.1016/S0140-6736(08)60695-9)
5. Bankir L, Bichet DG, Morgenthaler NG (2017) Vasopressin: physiology, assessment and osmosensation. *J Intern Med* 282:284–297. <https://doi.org/10.1111/joim.12645>
6. Thibonnier M, Coles P, Thibonnier A, Shoham M (2002) Molecular pharmacology and modeling of vasopressin receptors. *Prog Brain Res* 139:179–196. [https://doi.org/10.1016/s0079-6123\(02\)39016-2](https://doi.org/10.1016/s0079-6123(02)39016-2)
7. Facciorusso A, Amoruso A, Neve V, et al (2014) Role of vaptans in the management of hydroelectrolytic imbalance in liver cirrhosis. *World J Hepatol* 6:793–799. <https://doi.org/10.4254/wjh.v6.i11.793>
8. Aihara M, Fujiki H, Mizuguchi H, et al (2014) Tolvaptan delays the onset of end-stage renal disease in a polycystic kidney disease model by suppressing the increases in kidney volume and renal injury. *J Pharmacol Exp Ther*. <https://doi.org/10.1124/jpet.114.213256>
9. Boertien WE, Meijer E, de Jong PE, et al (2015) Short-term Effects of Tolvaptan in Individuals With Autosomal Dominant Polycystic Kidney Disease at Various Levels of Kidney Function. *Am J Kidney Dis Off J Natl Kidney Found*. <https://doi.org/10.1053/j.ajkd.2014.11.010>
10. Kelsey R (2013) Polycystic kidney disease: Tolvaptan in ADPKD-TEMPO 3:4 trial results. *Nat Rev Nephrol* 9:1. <https://doi.org/10.1038/nrneph.2012.236>
11. Manning M, Stoev S, Chini B, et al (2008) Peptide and non-peptide agonists and antagonists for the vasopressin and oxytocin V1a, V1b, V2 and OT receptors: research tools and potential therapeutic agents. *Prog Brain Res* 170:473–512. [https://doi.org/10.1016/S0079-6123\(08\)00437-8](https://doi.org/10.1016/S0079-6123(08)00437-8)
12. Manning M, Misicka A, Olma A, et al (2012) Oxytocin and Vasopressin Agonists and Antagonists as Research Tools and Potential Therapeutics. *J Neuroendocrinol* 24:609–628. <https://doi.org/10.1111/j.1365-2826.2012.02303.x>
13. Rondon-Berrios H, Berl T (2016) Vasopressin receptor antagonists: Characteristics and clinical role. *Best Pract Res Clin Endocrinol Metab* 30:289–303. <https://doi.org/10.1016/j.beem.2016.02.004>
14. Yamamura Y, Ogawa H, Yamashita H, et al (1992) Characterization of a novel aquaretic agent, OPC-31260, as an orally effective, nonpeptide vasopressin V2 receptor antagonist. *Br J Pharmacol* 105:787–791. <https://doi.org/10.1111/j.1476-5381.1992.tb09058.x>

15. Ohnishi A, Orita Y, Okahara R, et al (1993) Potent aquaretic agent. A novel nonpeptide selective vasopressin 2 antagonist (OPC-31260) in men. *J Clin Invest* 92:2653–2659. <https://doi.org/10.1172/JCI116881>
16. Drugs@FDA: FDA-Approved Drugs. <https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>. Accessed 23 Sep 2020
17. Ranieri M, Di Mise A, Tamma G, Valenti G (2019) Vasopressin–aquaporin-2 pathway: recent advances in understanding water balance disorders. *F1000Research* 8:. <https://doi.org/10.12688/f1000research.16654.1>
18. El Boustany R (2018) Vasopressin and Diabetic Kidney Disease. *Ann Nutr Metab* 72:17–20. <https://doi.org/10.1159/000488124>
19. Izumi Y, Miura K, Iwao H (2014) Therapeutic potential of vasopressin-receptor antagonists in heart failure. *J Pharmacol Sci* 124:1–6. <https://doi.org/10.1254/jphs.13r13cp>
20. Gassanov N, Semmo N, Semmo M, et al (2011) Arginine vasopressin (AVP) and treatment with arginine vasopressin receptor antagonists (vaptans) in congestive heart failure, liver cirrhosis and syndrome of inappropriate antidiuretic hormone secretion (SIADH). *Eur J Clin Pharmacol* 67:333–346. <https://doi.org/10.1007/s00228-011-1006-7>
21. Bodor N, Gabanyi Z, Wong CK (1989) A new method for the estimation of partition coefficient. *J Am Chem Soc* 111:3783–3786. <https://doi.org/10.1021/ja00193a003>
22. Hasel W, Hendrickson TF, Still WC (1988) A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput Methodol* 1:103–116. [https://doi.org/10.1016/0898-5529\(88\)90015-2](https://doi.org/10.1016/0898-5529(88)90015-2)
23. Miller KJ (1990) Additivity methods in molecular polarizability. *J Am Chem Soc* 112:8533–8542. <https://doi.org/10.1021/ja00179a044>
24. Parthasarathi R, Subramanian V, Roy DR, Chattaraj PK (2004) Electrophilicity index as a possible descriptor of biological activity. *Bioorg Med Chem* 12:5533–5543. <https://doi.org/10.1016/j.bmc.2004.08.013>
25. Frisch M, Trucks G, Schlegel H, et al (2009) Gaussian 09, Revision A.1. Gaussian Inc Wallingford CT
26. Froimowitz M (1993) HyperChem: a software package for computational chemistry and molecular modeling. *BioTechniques*
27. McFarland J, Gans D (1995) Multivariate Data Analysis of Chemical and Biological Data. Cluster Significance Analysis. In: *Chemometrics Methods in Molecular Design*. VCH Publishers, Inc, New York, NY, USA, pp 295–308
28. Alvarez-Ginarte YM, Montero-Cabrera LA, García-de la Vega JM, et al (2013) Integration of ligand and structure-based virtual screening for identification of leading anabolic steroids. *J Steroid Biochem Mol Biol* 138:348–358. <https://doi.org/10.1016/j.jsbmb.2013.07.004>
29. Yu S, Tranchevent L, Liu X, et al (2012) Optimized Data Fusion for Kernel k-Means Clustering. *IEEE Trans Pattern Anal Mach Intell* 34:1031–1039. <https://doi.org/10.1109/TPAMI.2011.255>

30. Weiß CH (2007) StatSoft, Inc., Tulsa, OK.: STATISTICA, Version 8. *AStA Adv Stat Anal* 91:339–341. <https://doi.org/10.1007/s10182-007-0038-x>
31. Alvarez- Ginarte YM, Crespo R, Montero- Cabrera LA, et al (2005) A novel in-silico approach for QSAR Studies of Anabolic and Androgenic Activities in the 17 β -hydroxy-5 α -androstane Steroid Family. *QSAR Comb Sci* 24:218–226. <https://doi.org/10.1002/qsar.200430889>
32. Liu P, Long W (2009) Current mathematical methods used in QSAR/QSPR studies. *Int J Mol Sci* 10:1978–1998. <https://doi.org/10.3390/ijms10051978>
33. Pourbasheer E, Aalizadeh R, Ganjali MR, Norouzi P (2013) QSAR study of IKK β inhibitors by the genetic algorithm: multiple linear regressions. *Med Chem Res* 23:57–66. <https://doi.org/10.1007/s00044-013-0611-7>
34. Pourbasheer E, Ahmadpour S, Zare-Dorabei R, Nekoei M (2017) Quantitative structure activity relationship study of p38 α MAP kinase inhibitors. *Arab J Chem* 10:33–40. <https://doi.org/10.1016/j.arabjc.2013.05.009>
35. Tropsha A (2010) Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol Inform* 29:476–488. <https://doi.org/10.1002/minf.201000061>
36. Gramatica P (2020) Principles of QSAR Modeling: Comments and Suggestions From Personal Experience. *Int J Quant Struct-Prop Relatsh IJQSPR* 5:61–97. <https://doi.org/10.4018/IJQSPR.20200701.oa1>
37. Oliveira DB de, Gaudio AC (2001) BuildQSAR: A New Computer Program for QSAR Analysis. *Quant Struct-Act Relatsh* 19:599–601. [https://doi.org/10.1002/1521-3838\(200012\)19:6<599::AID-QSAR599>3.0.CO;2-B](https://doi.org/10.1002/1521-3838(200012)19:6<599::AID-QSAR599>3.0.CO;2-B)
38. Larkin MA, Blackshields G, Brown NP, et al (2007) Clustal W and Clustal X version 2.0. *Bioinforma Oxf Engl* 23:2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>
39. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* 47:393–402. <https://doi.org/10.1002/prot.10104>
40. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. *J Mol Graph* 14:33–38, 27–28. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)
41. Stansfeld PJ, Goose JE, Caffrey M, et al (2015) MemProtMD: Automated Insertion of Membrane Protein Structures into Explicit Lipid Membranes. *Struct Lond Engl* 1993 23:1350–1361. <https://doi.org/10.1016/j.str.2015.05.006>
42. Phillips JC, Braun R, Wang W, et al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781–1802. <https://doi.org/10.1002/jcc.20289>
43. Guixà-González R, Rodríguez-Espigares I, Ramírez-Anguita JM, et al (2014) MEMBPLUGIN: studying membrane complexity in VMD. *Bioinforma Oxf Engl* 30:1478–1480. <https://doi.org/10.1093/bioinformatics/btu037>
44. Morris GM, Huey R, Lindstrom W, et al (2009) AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem* 30:2785–2791. <https://doi.org/10.1002/jcc.21256>

45. Trott O, Olson AJ (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31:455–461. <https://doi.org/10.1002/jcc.21334>
46. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612. <https://doi.org/10.1002/jcc.20084>
47. Durrant JD, McCammon JA (2011) BINANA: A Novel Algorithm for Ligand-Binding Characterization. *J Mol Graph Model* 29:888–893. <https://doi.org/10.1016/j.jmgm.2011.01.004>
48. Best RB, Zhu X, Shim J, et al (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J Chem Theory Comput* 8:3257–3273. <https://doi.org/10.1021/ct300400x>
49. Jo S, Kim T, Iyer VG, Im W (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* 29:1859–1865. <https://doi.org/10.1002/jcc.20945>
50. Lee J, Cheng X, Swails JM, et al (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem Theory Comput* 12:405–413. <https://doi.org/10.1021/acs.jctc.5b00935>
51. Almlöf M, Carlsson J, Åqvist J (2007) Improving the Accuracy of the Linear Interaction Energy Method for Solvation Free Energies. *J Chem Theory Comput* 3:2162–2175. <https://doi.org/10.1021/ct700106b>
52. Miranda WE, Noskov SY, Valiente PA (2015) Improving the LIE Method for Binding Free Energy Calculations of Protein-Ligand Complexes. *J Chem Inf Model* 55:1867–1877. <https://doi.org/10.1021/acs.jcim.5b00012>
53. Hansson T, Marelus J, Åqvist J (1998) Ligand binding affinity prediction by linear interaction energy methods. *J Comput Aided Mol Des* 12:27–35. <https://doi.org/10.1023/A:1007930623000>
54. Yu S, Tranchevent L, Liu X, et al (2012) Optimized Data Fusion for Kernel k-Means Clustering. *IEEE Trans Pattern Anal Mach Intell* 34:1031–1039. <https://doi.org/10.1109/TPAMI.2011.255>
55. Schauerl M, Podewitz M, Waldner BJ, Liedl KR (2016) Enthalpic and Entropic Contributions to Hydrophobicity. *J Chem Theory Comput* 12:4600–4610. <https://doi.org/10.1021/acs.jctc.6b00422>
56. Riley KE, Hobza P (2013) On the Importance and Origin of Aromatic Interactions in Chemistry and Biodisciplines. *Acc Chem Res* 46:927–936. <https://doi.org/10.1021/ar300083h>
57. Mignon P, Loverix S, Steyaert J, Geerlings P (2005) Influence of the π - π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res* 33:1779–1789. <https://doi.org/10.1093/nar/gki317>
58. Roy K, Kar S, Das RN (2015) Chapter 7 - Validation of QSAR Models. In: Roy K, Kar S, Das RN (eds) *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Academic Press, Boston, pp 231–289

59. Kučerka N, Nieh M-P, Katsaras J (2011) Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochim Biophys Acta* 1808:2761–2771. <https://doi.org/10.1016/j.bbamem.2011.07.022>
60. von Heijne G (1991) Proline kinks in transmembrane α -helices. *J Mol Biol* 218:499–503. [https://doi.org/10.1016/0022-2836\(91\)90695-3](https://doi.org/10.1016/0022-2836(91)90695-3)
61. Saleh N, Saladino G, Gervasio FL, et al (2016) A Three-Site Mechanism for Agonist/Antagonist Selective Binding to Vasopressin Receptors. *Angew Chem Int Ed Engl* 55:8008–8012. <https://doi.org/10.1002/anie.201602729>
62. Crombie AL, Antrilli TM, Campbell BA, et al (2010) Synthesis and evaluation of azabicyclo[3.2.1]octane derivatives as potent mixed vasopressin antagonists. *Bioorg Med Chem Lett* 20:3742–3745. <https://doi.org/10.1016/j.bmcl.2010.04.068>
63. Yamamura Y, Nakamura S, Itoh S, et al (1998) OPC-41061, a highly potent human vasopressin V2-receptor antagonist: pharmacological profile and aquaretic effect by single and multiple oral dosing in rats. *J Pharmacol Exp Ther* 287:860–867
64. Mouillac B, Chini B, Balestre MN, et al (1995) The binding site of neuropeptide vasopressin V1a receptor. Evidence for a major localization within transmembrane regions. *J Biol Chem* 270:25771–25777. <https://doi.org/10.1074/jbc.270.43.25771>
65. Albertazzi E, Zanchetta D, Barbier P, et al (2000) Nephrogenic Diabetes Insipidus: Functional Analysis of New AVPR2 Mutations Identified in Italian Families. *J Am Soc Nephrol* 11:1033–1043
66. Pasel K, Schulz A, Timmermann K, et al (2000) Functional characterization of the molecular defects causing nephrogenic diabetes insipidus in eight families. *J Clin Endocrinol Metab* 85:1703–1710. <https://doi.org/10.1210/jcem.85.4.6507>
67. Makita N, Manaka K, Sato J, Iiri T (2020) V2 vasopressin receptor mutations. *Vitam Horm* 113:79–99. <https://doi.org/10.1016/bs.vh.2019.08.012>
68. Cotte N, Balestre M-N, Aumelas A, et al (2000) Conserved aromatic residues in the transmembrane region VI of the V1a vasopressin receptor differentiate agonist vs. antagonist ligand binding. *Eur J Biochem* 267:4253–4263. <https://doi.org/10.1046/j.1432-1033.2000.01472.x>
69. Neocleous V, Skordis N, Shammas C, et al (2012) Identification and characterization of a novel X-linked AVPR2 mutation causing partial nephrogenic diabetes insipidus: A case report and review of the literature. *Metabolism* 61:922–930. <https://doi.org/10.1016/j.metabol.2012.01.005>
70. Sahakitrungruang T, Tee MK, Rattanachartnarong N, et al (2010) Functional characterization of vasopressin receptor 2 mutations causing partial and complete congenital nephrogenic diabetes insipidus in Thai families. *Horm Res Pædiatrics* 73:349–354. <https://doi.org/10.1159/000308167>