The Signaling Pathway of Dopamine D2 Receptor (D2R) Activation Using Normal Mode Analysis (NMA) and the Construction of Universal Pharmacophore Models for D2R Inhibitors

Ramin Ekhteiari Salmas[1*], Matthias Stein[2], Mine Yurtsever[3], Philip Seeman[4], Mert Mestanoglu[1], Ismail Erol[1], Serdar Durdagi[1*]

[1]Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey
[2]Max-Planck Institute for Dynamics of Complex Technical System, Molecular Simulations and Design Group, Sandtorstrasse 1, 39106, Magdeburg, Germany
[3]Department of Chemistry, Istanbul Technical University, Istanbul, Turkey
[4]Departments of Pharmacology and Psychiatry, University of Toronto, 260 Heath Street West, Unit 605, M5P 3L6, Toronto, Ontario, Canada

Corresponding Authors:

*Assoc. Prof. Dr. Serdar Durdagi, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey. E-mail: serdar.durdagi@bahcesehir.edu.tr

*Ramin Ekhteiari Salmas, Ph.D, Department of Biophysics, School of Medicine, Bahcesehir University, Istanbul, Turkey. E-mail: ekhteiarisalmas@itu.edu.tr

Abstract

G-protein-coupled receptors (GPCRs) are targets of more than 30% of marketed drugs. Investigation on the GPCRs may shed light on upcoming drug design studies. In the present study, we performed a combination of receptor- and ligand-based analysis targeting the dopamine D2 receptor (D2R). The signaling pathway of D2R activation and the construction of universal pharmacophore models for D2R inhibitors were studied. The key amino acids, which contributed to the regular activation of the D2R, were in detail investigated by means of Normal Mode Analysis (NMA). A derived cross-correlation matrix provided us a an understanding of the degree of pair residue correlations. Although negative correlations were not observed in the case of the inactive D2R state, a high degree of correlation appeared between the residues in the active state. NMA results showed that the cytoplasmic side of the TM5 plays a significant role in promoting of residue-residue correlations in the active state of D2R. Tracing motions of the amino acids Arg219, Arg220, Val223, Asn224, Lys226 and Ser228 in the position of the TM5 are found to be critical in signal transduction. Complementing the receptor-based modeling, ligand-based modeling is also performed using known D2R inhibitors. The top-scored pharmacophore model was found as 5-sited (AADPR.671) hypothesis from PHASE modeling from a pool consisting of more than hundred initial candidates. The constructed model using 38 D2R inhibitors (in the training set) was validated with 15 additional test set compounds. The resulting model correctly predicted the pIC_{50} values of an additional test set compounds as true unknowns.

Introduction

G-protein-coupled receptors (GPCRs) constitute the largest membrane-bound protein classes [Römpler, H.; Stäubert, C.; Thor, D.; Schulz, A.; Hofreiter, M.; Schöneberg, T. G Protein-Coupled Time Travel: Evolutionary Aspects of GPCR Research. Mol. Interv. 2007, 7 (1), 17–25.]. These proteins play a pivotal role in cell signaling processes in response to neurotransmitters as well as hormones [Congreve, M.; Marshall, F. The Impact of GPCR Structures on Pharmacology and Structure-Based Drug Design. Br. J. Pharmacol. 2010, 159 (5), 986–996.]. Since the human GPCRs are the main targets of the majority of marketed drugs, focusing on their inhibition mechanism can significantly develop our insight on drug design studies [Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? Nat. Rev. Drug Discov. 2006, 5 (12), 993–996.] . The GPCRs shuttle between active and inactive states upon drug binding. Antipsychotic drugs are mostly used in schizophrenia therapies and they may partly inhibit the target structure in an inactive state of the receptor [Laruelle, M.; Abi-Dargham, a; van Dyck, C. H.; Gil, R.; D'Souza, C. D.; Erdos, J.; McCance, E.; Rosenblatt, W.; Fingado, C.; Zoghbi, S. S.; Baldwin, R. M.; Seibyl, J. P.; Krystal, J. H.; Charney, D. S.; Innis, R. B. Single Photon Emission Computerized Tomography Imaging of Amphetamine-Induced Dopamine Release in Drug-Free Schizophrenic Subjects. Proc. Natl. Acad. Sci. U. S. A. 1996, 93 (17), 9235–9240.]. There are around 390 GPCRs encoded in the human genome, which are categorized into 6 classes based on sequence homology and functional similarities: Class A (Rhodopsin-like), Class B (Secretin receptor family), Class C (Metabotropic glutamate), Class D (Fungal mating pheromone receptors), Class E (Cyclic AMP receptors) and Class F (Frizzled/Smoothened) [Lagerström, M. C.; Schiöth, H. B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. Nat. Rev. Drug Discov. 2008, 7 (4), 339–357.-Foord, S. M.; Bonner, T. I.; Neubig, R. R.; Rosser, E. M.; Pin, J.; Davenport, A. P.; Spedding, M.; Harmar, A. J. International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List. Pharmacol. Rev. 2005, 57 (2), 279–288.]. GPCR crystal structures were mainly obtained in their inactive forms. Active-state structures (i.e., partial or full active) are only available

for β2-adrenergic receptor (β2AR) [Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. a; Lyons, J. a; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the β2 Adrenergic Receptor-Gs Protein Complex. Nature 2011, 477 (7366), 549–555.], adenosine 2A receptor (A2AR) [Ijzerman, A. P.; Jaakola, V.-P.; Cherezov, V.; Stevens, R. C.; Griffith, M. T.; Lane, J. R.; Hanson, M. A.; Chien, E. Y. T. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. Science 2008, 322, 1211-1217.], muscarinic M2 receptor (M2R) [Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the Human M2 Muscarinic Acetylcholine Receptor Bound to an Antagonist. Nature 2012, 482 (7386), 547–551.], and rhodopsin (RHO) [Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H.-W.; Ernst, O. P. Crystal Structure of the Ligand-Free G-Protein-Coupled Receptor Opsin. Nature 2008, 454 (7201), 183–187.,Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P. Crystal Structure of Opsin in Its G-Protein-Interacting Conformation. Nature 2008, 455 (7212), 497–502.].

The current study focuses on the Dopamine D2 Receptor (D2R) that plays a significant role in the nervous system, and in a variety of disorders in the central nervous system (CNS), such as Parkinson's disease and schizophrenia where dysfunction occurs in the regular signal transfer from the D2R [Seeman, P.; Lee, T.; Chau-Wong, M.; Wong, K. Antipsychotic Drug Doses and Neuroleptic/dopamine Receptors. Nature 1976, 261, 717–719.,Creese, I.; Burt, D. R.; Snyder, S. H. Dopamine Receptor Binding Predicts Clinical and Pharmacological Potencies of Antischizophrenic Drugs. J. Neuropsychiatry Clin. Neurosci. 1996, 8 (2), 223–226.]. Since high-resolution crystal structures of D2R are not available yet, most of the studies of dopaminergic systems are performed using *in silico*-generated 3D model structures of D2R. In order to investigate drug screening as well as conformational transitions of D2Rs, previously derived 3D model structures of active and inactive states of D2Rs by our group were used [Salmas, R. E.; Yurtsever, M.; Stein, M.; Durdagi,

S. Modeling and Protein Engineering Studies of Active and Inactive States of Human Dopamine D2 Receptor (D2R) and Investigation of Drug/receptor Interactions. Mol. Divers. 2015, 19 (2), 321–332.-Salmas RE, Yurtsever M, Durdagi S. Atomistic molecular dynamics simulations of typical and atypical antipsychotic drugs at the dopamine D2 receptor (D2R) elucidates their inhibition mechanism, J Biomol Struct Dyn. 2016 Apr 6:1-17. [In press]. DOI: 10.1080/07391102.2016.1159986.]. The alignment of the active and inactive D2R models (Figure S1, Supplementary Materials) was performed based on the alpha carbon atoms, in which the conformational changes of the protein while activation are monitored. The signaling pathway of the full active D2R structure is investigated in this study.

It is obvious that all marketed antipsychotics produce their effect through the antagonizing of D2R has proved that enhanced D2R signaling is a major part of the pathophysiology of schizophrenia [Seeman, P. (2006). Targeting the dopamine D2 receptor in schizophrenia. Expert Opinion on Therapeutic Targets, 10(4), 515–531. doi:10.1517/14728222.10.4.515,Seeman, P., & Lee, T. (1975). Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. Science (New York, N.Y.), 188(4194), 1217–1219. doi:10.1126/science.1145194] . The internal motions and allosteric regulations of GPCRs have critically been identified as essential for signal transferring. For this aim, Normal Mode Analysis (NMA), described by the normal modes appeared at the fixed frequencies, [Case, D. A. Normal Mode Analysis of Protein Dynamics. Curr. Opin. Struct. Biol. 1994, 4 (2), 285–290.] was implemented in order to identify the amino acids that are mainly participating in a conformational transition upon D2R activation. In the last two decades, there have been many investigations using NMA on different proteins [Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. Chemical Reviews, 110(3), 1463–1497. doi:10.1021/cr900095e-Su, P. C., Tsai, C. C., Mehboob, S., Hevener, K. E., & Johnson, M. E. (2015). Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI). Journal of

Computational Chemistry, 36(25), 1859–1873. doi:10.1002/jcc.24011]. These studies have proven this method as useful in predicting the collective modes, especially for the motions at the low frequencies [Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. Chemical Reviews, 110(3), 1463–1497. doi:10.1021/cr900095e]. An earlier use of NMA to predict the internal motion of a protein is attributed to Karplus *et al.* (1983) [Brooks, B., Brooks, B., Karplus, M., & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A, 80(21), 6571–6575. doi:10.1073/pnas.80.21.6571]. They performed NMA employing of an empirical potential energy function comprising bonded and non-bonded interactions for the bovine pancreatic trypsin inhibitor [Brooks, B., Brooks, B., Karplus, M., & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A, 80(21), 6571–6575. doi:10.1073/pnas.80.21.6571]. Their finds suggested that, in spite of the anharmonic contributions to the potential, a normal mode calculation can reveal useful information corresponding to the internal motions of proteins [Brooks, B., Brooks, B., Karplus, M., & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci U S A, 80(21), 6571–6575. doi:10.1073/pnas.80.21.6571]. Another considerable effort was performed by Skjaerven et al [Skjaerven, L., Martinez, A., & Reuter, N. (2011). Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. Proteins, 79(1), 232–43. doi:10.1002/prot.22875] toward studying of internal motion of a subunit of the GroEL chaperone, using NMA.

In this study, not just active and inactive modes of receptor-based models of D2R, also construction of a ligand-based pharmacophore modeling is also performed for better understanding of D2R system. A three-dimensional quantitative structure–activity relationships (3D-QSAR) study may assist to reveal which features of a molecule result to its entire activity and derived information can be used to discover new molecules with modified properties.

Methods

Normal Mode Analysis (NMA)

The vast variety of simulation methods available makes it difficult to select a most appropriate way to employ in studying of the living systems. Many factors should be taken into account when choosing the optimal protocol. Implementation of molecular dynamics (MD) simulations and Quantum Mechanics (QM) methods are too expensive when the dynamics of large biological systems are considered. In such cases, it is not practical to contemplate monitoring completely all the essential displacements of the atoms using MD simulations. Rather, Normal Mode Analysis (NMA) which is one of the commonly used time-independent types of simulations, that searches long and local conformational changes in protein structures, can be used [Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. Chemical Reviews, 110(3), 1463–1497. doi:10.1021/cr900095e]. NMA as a low-time consuming approach uses a statistical approach to monitor the vibrational motion and flexibility of protein atoms between pre-existing conformations. Normal modes are useful as they are responsible for collective motions of the particles in an ensemble [Leach, A. R. (2001). Molecular Modelling: Principles and Applications. AddisonWesley Longman Ltd.]. The number of trajectories used is based on the norm modes available at the NMA (which was around 50 in this study). The frequencies of the normal modes together with the shifting of the particles can be estimated using a force field method or wave function of the Hessian matrix of second derivatives [Leach, A. R. (2001). Molecular Modelling: Principles and Applications. AddisonWesley Longman Ltd.]. In order to simplify the calculations, Cα atoms of the protein were only considered in this study. NMA was performed for the active and inactive models of the D2R using Bio3D [Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. Bioinformatics 2006, 22 (21), 2695–2696.]. Herein, the second derivative matrix of the potential energy was calculated, *V*. The normal mode vectors were obtained by the eigenvalue equitation (Equation 1).

$$A^T V A = \lambda \qquad (1)$$

The normal modes are described by eigenvectors (A) and their eigenvalues (λ). The eigenvalues were connected to a conformer transition of the protein along the specified eigenvectors.

Three-dimensional Quantitative Structure-Activity Relationships (3D-QSAR)

The 3D-QSAR approach was performed using PHASE module of Maestro molecular modeling package [Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. a. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. J. Comput. Aided. Mol. Des. 2006, 20 (10-11), 647–671.]. Fifty-three approved D2R inhibitors were retrieved from the ZINC database [Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. J. Chem. Inf. Model. 2012, 52 (7), 1757–1768.] (Table S1, Supplementary Materials). Activity of the each ligand against the D2R was considered, based on their corresponding IC_{50} values (pIC_{50} values were used in the calculations). All chemical structures of the ligands were prepared (protonation states are determined with PROPKA) and energy minimization studies were performed at the physiological pH (pH = 7.4) using ligand preparation module (LigPrep) of Maestro molecular modeling package [Schrödinger Release 2015-4: LigPrep, version 3.6, Schrödinger, LLC, New York, NY, 2015.] with the OPLS 2005 forcefield [Siu, S. W. I.; Pluhackova, K.; Böckmann, R. a. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. J. Chem. Theory Comput. 2012, 8 (4), 1459–1470.]. Not just one conformer of the molecules at the database was used for the simulations; instead a conformational search was set up using a Monte Carlo Multiple Minimum (MCMM) method [Schrödinger Release 2015-4: MacroModel, version 11.0, Schrödinger, LLC, New York, NY, 2015.]. The cut-off energy value for saving the conformers was adjusted to 21.0 kJ/mol. A RMSD cutoff of 1 Å was employed to remove closely similar conformers. Ligands that have diverse activity against D2R were used. PHASE utilizes an analysis of *k*-point pharmacophores derived from a conformational set of the active ligands and then determines all spatial arrangements with pharmacophore features shared by

those ligands. The functional interactions between the receptor and the ligand rely on conformational complementarity between main groups placed in the ligand and orientation of amino acid residues in the active site of the target structure. PHASE identifies these main interactions as hydrogen bond acceptors (labeled as 'A'), hydrogen bond donors (labeled as 'D'), hydrophobic groups (labeled as 'H'), negatively charged groups (labeled as 'N'), positively charged groups (labeled as 'P'), and aromatic rings (labeled as 'R').

Results and Discussion

Conformational Stability

The biochemical and biophysical properties of a protein are determined mostly upon the conformations that it can adopt. The critical feature of the conformational analysis is the flexibility calculations of the atoms. Herein, in order to better understand the local and the global conformational changes of the active and inactive models of D2R, a flexibility assay was performed for both systems based on their normal modes. Atomistic fluctuations of individual amino acids were calculated for the models, as shown in Figure 1. The residual flexibility graphs illustrated the amplitude of motion of the each residue for each corresponding mode. It was observed that the terminal residues and the loop domains revealed high fluctuation values for both the active and inactive states and low flexibility were observed for the trans-membrane (TM) alpha-helical domains, which is in good agreement with experimental suggestions (based on the B factors or the temperature factors of the both active and inactive states of the β2AR) [Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. a; Lyons, J. a; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the β2 Adrenergic Receptor-Gs Protein Complex. Nature 2011, 477 (7366), 549–555.]. Comparative stability analysis between two conformers indicated a higher structural flexibility in the cytoplasmic half of the TM5 and TM6 for the active form compared with the inactive form. (Figure 1) Indeed, a higher mobility for the active conformer is expected. Particularly, the intra-cellular loops in the active state

exhibited higher degree of flexibility compared to the inactive state due to the conformational transition at the cytoplasmic side. The third intra-cellular loop in the active form, connecting the TM5 to the TM6 in the cytoplasmic side, has the ability to link to the G-protein in signal transduction processes. In order to clearly map the position and degree of the motion over the models, superimposition of the trajectory frames obtained by the NMA was studied as shown in Figure 2. The comparative conformational analysis can clearly monitor the degree of dynamical activity on the structures. This approach successfully illustrates the main differences in the mobile portions of the models and monitors the influence of the conformational transition on the motion of the D2R. The effect of the biological activity on the conformation was highly pronounced on the cytoplasmic half of the active form, whereas, these domain did not revealed significant deviation in the inactive state (Figure 2). By proceeding in this fashion through the conformations available to each model, we could further enhance the efficiency of this analysis by calculating of the root-mean-square deviations (RMSDs) of the individual models respect to the reference structure. RMSDs of 1.56 and 0.95 Å for the active and inactive states were observed, respectively. These results are expected and the larger RMSD value is due to the distinct deviations of the Cα atoms in the active form.

The Signaling Pathway of the D2R

In order to probe the signal transduction pathway of D2R, which follows the D2R upon the regular activation, we performed a residue-based interaction correlation analysis. The prevalence of both full active and inactive D2R protein structural models encouraged us to shed light on the signal transduction of the D2Rs and identify the key amino acids contributing to the regular activation by monitoring the correlation degree between the residues. The correlation maps for the TMs of D2R models were studied as shown in Figure 3. The cross-correlation matrix was mapped for the each model based on the NMA. Black boxes and blank domain represent alpha helices and loop domain, respectively in Figure 3. Colors which varying from blue (positive correlation) to red (negative correlation) are identifying the intensity of correlation for the each residue. (Figure 3) The

correlation profiles along both axes show the amino acid positions and corresponding secondary structure data. The matrix provides us a better understanding of the degree of pair residue correlations, which were taking place in the protein (the domains are labeled in Figure 3). Although negative correlations were not observed in the case of the inactive D2R state, a high degree of the correlation appeared between the residues in the active state. (Figure 3) The correlations are mainly associated with the conformational transition, which was occurring at the cytoplasmic half of the active state. In addition, in order to monitor the critical residues at the D2R active state which participated in the correlations, these pair residues are represented as blue balls in Figure 4. The correlation between the main residues at the active form is illustrated with dashed lines. (Figure 4) Judging from the profile we found that all the correlations were established at the cytoplasmic half of the TM5, TM6, TM7 and TM8 domains. The cytoplasmic side of the TM5 plays a significant role in promoting of these correlations. Arg219, Arg220, Val223, Asn224, Lys226 and Ser228 amino acids in the position of the TM5 are the critical amino acid in signal transduction, which were establishing correlations with the specific amino acid residues in the TM6 (Ile377, Leu379, Gly380, Val380, Phe382, Ile383 and Ile384), TM7 (Val421, Ile424, Ile425, Tyr426, Thr427, Thr428, Phe429) and TM8 (Asn430, Ile431, Glu432, Phe433, Arg434, Lys435, Ala436 and Phe437). The predicted results are in a good agreement with experiments, as the critical role of the amino acids located in the TM5, TM6 and TM7 for signal transfer from outside the cell through the D2R is reported by Javitch *et al* [Javitch, J. A., Ballesteros, J. A., Weinstein, H., & Chen, J. (1998). A cluster of aromatic residues in the sixth membrane-spanning segment of the dopamine D2 receptor is accessible in the binding-site crevice. Biochemistry, 37(4), 998–1006. doi:10.1021/bi972241y-Javitch, J. A. (1995). Residues in the Fifth Membrane-Spanning Segment of the Dopamine, 16433–16439.].

3D QSAR Approach

3D QSAR is a useful method for the construction of universal pharmacophore models for target structures. It provides important data particularly when there is no crystallographic information for

a protein. In order to predict and construct common pharmacophores of the D2R inhibitors, fifty three approved D2R inhibitors were incorporated in the 3D QSAR simulations (Figure S2, Supplementary Materials). A conformational search was performed for each individual ligand in the database using a Monte Carlo method before the 3D QSAR simulations. The pIC_{50} values of compounds in database have a diverse activity ranged between $5.27 - 10.00$. The pIC_{50} values of highest and the lowest active compound were associated with rotigotine (ZINC00004028) and 1-Benzyl-4-(2-iodo-1H-pyrrol-1-yl) piperidine (ZINC20358060), respectively. To apply an activity threshold for ligands, pIC_{50} threshold was set as $pIC_{50} > 7.2$ and $pIC_{50} < 6.2$ for the active and inactive compounds, respectively [Zhu, T., Cao, S., Su, P.-C., Patel, R., Shah, D., Chokshi, H. B., … Hevener, K. E. (2013). Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. Journal of Medicinal Chemistry, 56(17), 6560–72. doi:10.1021/jm301916b]. Pharmacophore features were used to derive pharmacophore sites for all the compounds. PHASE defines six pharmacophore sites as follows; hydrogen bond acceptor ("A"), hydrogen bond donor ("D"), hydrophobic group ("H"), negatively charged group ("N"), positively charged group ("P") and aromatic ring ("R"). The number of sites for each pharmacophore was investigated between 5 to 7. Common pharmacophore motives found for D2R inhibitors include; AADPR, AADRR, AAHPR, AAPRR, ADHRR, ADPRR, AHHPR, AHHRR, AHPRR and HHPRR, and were employed in the subsequent 3D QSAR modeling steps. The regression was applied using a partial least squares (PLS) method, in which sets of models were generated with an increasing number of PLS factors. In order to validate the derived models, all the collected compounds were divided into training (38 compounds) and test sets (15 compounds). Statistical validation results and derived scores for the top-ranked 3D QSAR hypotheses (AADPR.671, AAPRR.3900, AADRR.1398, and ADHRR.2864) are listed in Table 1. Following parameters were considered to rank and choose the pharmacophore models (highest R^2, Q^2 and Pearson R and lowest RMSE, SD, and P values). The correlation of experimental and predicted pIC_{50} values for the training and test set molecules are plotted in Figure 5. The correlation

coefficients for training set and test set compounds are found as 0.95 and 0.75 respectively. The high correlation coefficient ($R^2 = 0.95$) for this generated pharmacophore model suggests the reliability of the 3D QSAR model. Figure 6 represents the alignment of the top pharmacophore model AADPR.671 with its reference compound (template) (ZINC11909395), as well as selected active (ZINC00538312/risperidone) and inactive ligands (ZINC10450796). Risperidone was introduced as a highly active ligand, which has a high alignment score with the constructed model. However, compound ZINC10450796 as a highly inactive ligand reveals a low alignment degree with the pharmacophore model. The resulting model correctly predicted the pIC_{50} values of test set compounds as true unknowns, hence these results validate the reliability of the constructed 3D QSAR model. The 3D pharmacophore model has the potential to be used for subsequent ligand-based virtual screening studies to discover new D2R inhibitor candidates.

Conclusions

In summary, two distinct approaches were used (i) NMA for the D2R signaling pathway; (ii) ligand-based modeling for development of an universal pharmacophore model for D2R inhibition. The NMA was implemented to identify the main amino acids at the cytoplasmic half of the active model and to monitor their structural correlations, which were taking place between the TMs. Upon GPCR activation, the TM helices (i.e., TM5, TM6 and TM7) undergo a large correlated motion initiated by conformational changes at the cytoplasmic side. In addition, a conformational stability analysis was performed for both models using the NMA method. As expected, the active D2R form was more flexible and mobile based on the structural results compared to the inactive D2R model. 38 D2R inhibitors are used as training set, and a new 3D QSAR model was generated to suggest a common and popular pharmacophore hypothesis for the D2R inhibitors which is tested by 15 internal test set molecules. In our work, a five pharmacophore sited model with site types of AADPR was found to be the most prominent ligand-based interaction model which can be used as an universal pharmacophore model for subsequent D2R ligand-based screening.

Acknowledgements

References

1. Römpler, H.; Stäubert, C.; Thor, D.; Schulz, A.; Hofreiter, M.; Schöneberg, T. G Protein-Coupled Time Travel: Evolutionary Aspects of GPCR Research. *Mol. Interv.* **2007**, *7* (1), 17–25.
2. Congreve, M.; Marshall, F. The Impact of GPCR Structures on Pharmacology and Structure-Based Drug Design. *Br. J. Pharmacol.* **2010**, *159* (5), 986–996.
3. Overington, J. P.; Al-Lazikani, B.; Hopkins, A. L. How Many Drug Targets Are There? *Nat. Rev. Drug Discov.* **2006**, *5* (12), 993–996.
4. Laruelle, M.; Abi-Dargham, a; van Dyck, C. H.; Gil, R.; D'Souza, C. D.; Erdos, J.; McCance, E.; Rosenblatt, W.; Fingado, C.; Zoghbi, S. S.; Baldwin, R. M.; Seibyl, J. P.; Krystal, J. H.; Charney, D. S.; Innis, R. B. Single Photon Emission Computerized Tomography Imaging of Amphetamine-Induced Dopamine Release in Drug-Free Schizophrenic Subjects. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (17), 9235–9240.
5. Lagerström, M. C.; Schiöth, H. B. Structural Diversity of G Protein-Coupled Receptors and Significance for Drug Discovery. *Nat. Rev. Drug Discov.* **2008**, *7* (4), 339–357.
6. Attwood, T. K.; Findlay, J. B. Fingerprinting G-Protein-Coupled Receptors. *Protein Eng.* **1994**, *7* (2), 195–203.
7. Kolakowski, L. F. GCRDb: A G-Protein-Coupled Receptor Database. *Receptors Channels* **1994**, *2* (1), 1–7.
8. Foord, S. M.; Bonner, T. I.; Neubig, R. R.; Rosser, E. M.; Pin, J.; Davenport, A. P.; Spedding, M.; Harmar, A. J. International Union of Pharmacology. XLVI. G Protein-Coupled Receptor List. *Pharmacol. Rev.* **2005**, *57* (2), 279–288.
9. Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. a; Lyons, J. a; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal Structure of the β2 Adrenergic Receptor-Gs Protein Complex. *Nature* **2011**, *477* (7366), 549–555.
10. Ijzerman, A. P.; Jaakola, V.-P.; Cherezov, V.; Stevens, R. C.; Griffith, M. T.; Lane, J. R.; Hanson, M. A.; Chien, E. Y. T. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322,* 1211-1217.
11. Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the Human M2 Muscarinic Acetylcholine Receptor Bound to an Antagonist. *Nature* **2012**, *482* (7386), 547–551.
12. Park, J. H.; Scheerer, P.; Hofmann, K. P.; Choe, H.-W.; Ernst, O. P. Crystal Structure of the Ligand-Free G-Protein-Coupled Receptor Opsin. *Nature* **2008**, *454* (7201), 183–187.

13. Scheerer, P.; Park, J. H.; Hildebrand, P. W.; Kim, Y. J.; Krauss, N.; Choe, H. W.; Hofmann, K. P.; Ernst, O. P. Crystal Structure of Opsin in Its G-Protein-Interacting Conformation. *Nature* **2008**, *455* (7212), 497–502.

14. Seeman, P.; Lee, T.; Chau-Wong, M.; Wong, K. Antipsychotic Drug Doses and Neuroleptic/dopamine Receptors. *Nature* **1976**, *261*, 717–719.

15. Creese, I.; Burt, D. R.; Snyder, S. H. Dopamine Receptor Binding Predicts Clinical and Pharmacological Potencies of Antischizophrenic Drugs. *J. Neuropsychiatry Clin. Neurosci.* **1996**, *8* (2), 223–226.

16. Salmas, R. E.; Yurtsever, M.; Stein, M.; Durdagi, S. Modeling and Protein Engineering Studies of Active and Inactive States of Human Dopamine D2 Receptor (D2R) and Investigation of Drug/receptor Interactions. *Mol. Divers.* **2015**, *19* (2), 321–332.

17. Durdagi, S.; Salmas, R. E.; Stein, M.; Yurtsever, M.; Seeman, P. Binding Interactions of Dopamine and Apomorphine in D2High and D2Low States of Human Dopamine D2 Receptor (D2R) Using Computational and Experimental Techniques. *ACS Chem. Neurosci.* **2015**, acschemneuro.5b00271.

18. Salmas RE, Yurtsever M, Durdagi S. Atomistic molecular dynamics simulations of typical and atypical antipsychotic drugs at the dopamine D2 receptor (D2R) elucidates their inhibition mechanism, J Biomol Struct Dyn. 2016 Apr 6:1-17. [In press]. DOI: 10.1080/07391102.2016.1159986.

19. Seeman, P. (2006). Targeting the dopamine D2 receptor in schizophrenia. *Expert Opinion on Therapeutic Targets*, *10*(4), 515–531. doi:10.1517/14728222.10.4.515

20. Seeman, P., & Lee, T. (1975). Antipsychotic drugs: direct correlation between clinical potency and presynaptic action on dopamine neurons. *Science (New York, N.Y.)*, *188*(4194), 1217–1219. doi:10.1126/science.1145194

21. Case, D. A. Normal Mode Analysis of Protein Dynamics. *Curr. Opin. Struct. Biol.* **1994**, *4* (2), 285–290.

22. Bahar, I., Lezon, T. R., Bakan, A., & Shrivastava, I. H. (2010). Normal mode analysis of biomolecular structures: Functional mechanisms of membrane proteins. *Chemical Reviews*, *110*(3), 1463–1497. doi:10.1021/cr900095e

23. Miloshevsky, G. V., & Jordan, P. C. (2006). The Open State Gating Mechanism of Gramicidin A Requires Relative Opposed Monomer Rotation and Simultaneous Lateral Displacement. *Structure*, *14*(8), 1241–1249. doi:10.1016/j.str.2006.06.007

24. Bahar, I., & Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struc. Biol.*, *15*(5), 586–592. doi:10.1016/j.sbi.2005.08.007

25. Cui, Q., & Bahar, I. (2005). *Normal Mode Analysis: Theory and Applications to Biological and Chemical Systems. Chapman and Hall/CRC*. Retrieved from http://www.crcpress.com/product/isbn/9781584884729

26. Su, P. C., Tsai, C. C., Mehboob, S., Hevener, K. E., & Johnson, M. E. (2015). Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of F. tularensis enoyl-ACP reductase (FabI). *Journal of Computational Chemistry*, *36*(25), 1859–1873. doi:10.1002/jcc.24011

27. Brooks, B., Brooks, B., Karplus, M., & Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A*, *80*(21), 6571–6575. doi:10.1073/pnas.80.21.6571

28. Skjaerven, L., Martinez, A., & Reuter, N. (2011). Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins*, *79*(1), 232–43. doi:10.1002/prot.22875

29. Leach, A. R. (2001). *Molecular Modelling: Principles and Applications. AddisonWesley Longman Ltd*.

30. Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695–2696.

31. Dixon, S. L.; Smondyrev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. a. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *J. Comput. Aided. Mol. Des.* **2006**, *20* (10-11), 647–671.

32. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52* (7), 1757–1768.

33. **Schrödinger Release 2015-4:** LigPrep, version 3.6, Schrödinger, LLC, New York, NY, 2015.

34. Siu, S. W. I.; Pluhackova, K.; Böckmann, R. a. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *J. Chem. Theory Comput.* **2012**, *8* (4), 1459–1470.

35. **Schrödinger Release 2015-4:** MacroModel, version 11.0, Schrödinger, LLC, New York, NY, 2015.

36. Javitch, J. A., Ballesteros, J. A., Weinstein, H., & Chen, J. (1998). A cluster of aromatic residues in the sixth membrane-spanning segment of the dopamine D2 receptor is accessible in the binding-site crevice. *Biochemistry*, *37*(4), 998–1006. doi:10.1021/bi972241y

37. Fu, D., Ballesteros, J. a, Weinstein, H., Chen, J., & Javitch, J. a. (1996). Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding-site crevice. *Biochemistry*, *35*(35), 11278–85. doi:10.1021/bi960928x

38. Javitch, J. A. (1995). Residues in the Fifth Membrane-Spanning Segment of the Dopamine, 16433–16439.

39. Zhu, T., Cao, S., Su, P.-C., Patel, R., Shah, D., Chokshi, H. B., … Hevener, K. E. (2013). Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *Journal of Medicinal Chemistry*, *56*(17), 6560–72. doi:10.1021/jm301916b

TABLES

Table 1. Statistical parameter values for the top 3D-QSAR hypotheses AADPR.671, AAPRR.3900, AADRR.1398, and ADHRR.2864.

Statistical parameter values for the AADPR.671								
PLS factor	SD	R^2	F	P	RMSE	Q^2	R-Pearson	Reference ligand
1	0.58	0.80	120.2	5.15×10^{-12}	0.66	0.61	0.78	ZINC11909395
2	0.30	0.95	268.5	1.95×10^{-19}	0.55	0.73	0.86	ZINC11909395
3	0.14	0.99	872.9	9.48×10^{-28}	0.55	0.73	0.87	ZINC11909395

Statistical parameter values for AAPRR.3900								
PLS factor	SD	R^2	F	P	RMSE	Q^2	R-Pearson	Reference ligand
1	0.72	0.68	66.7	3.16×10^{-09}	0.55	0.67	0.86	ZINC04364519
2	0.35	0.93	197.0	5.59×10^{-18}	0.63	0.56	0.82	ZINC04364519
3	0.22	0.97	344.5	9.15×10^{-23}	0.62	0.58	0.84	ZINC04364519

Statistical parameter values for AADRR.1398								
PLS factor	SD	R^2	F	P	RMSE	Q^2	R-Pearson	Reference ligand
1	0.52	0.82	118.3	3.57×10^{-11}	0.55	0.59	0.85	ZINC11909395
2	0.29	0.95	214.9	1.78×10^{-16}	0.53	0.62	0.86	ZINC11909395
3	0.16	0.98	504.5	8.38×10^{-22}	0.50	0.66	0.87	ZINC11909395

| Statistical parameter values for ADHRR.2864								
PLS factor	SD	R^2	F	P	RMSE	Q^2	R-Pearson	Reference ligand
1	0.75	0.66	54.0	5.29×10^{-08}	0.68	0.55	0.76	ZINC11909395
2	0.42	0.90	119.6	3.82×10^{-12}	0.58	0.67	0.85	ZINC11909395
3	0.20	0.98	370.1	1.95×10^{-21}	0.57	0.68	0.87	ZINC11909395

Figure 1. Flexibility graph for the each residue, calculated by normal mode analyses (NMA) for inactive and active D2R protein 3D structures. The boxes and blanks represent alpha helices and loop domains, respectively.
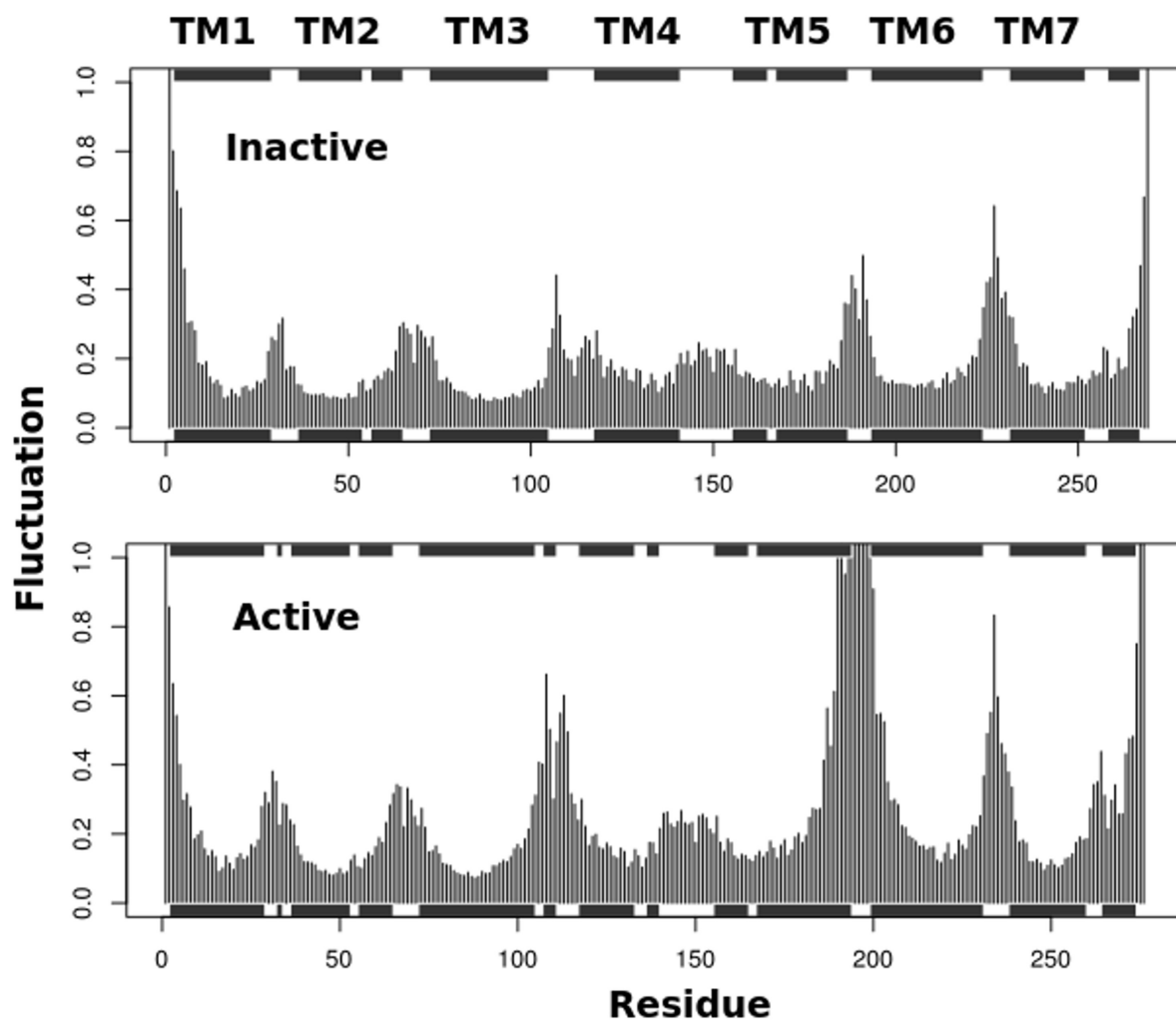
Figure 2. Trajectory frames, which were generated based on NMA. The active state revealed a high degree of structural mobility for all domains compared with the inactive form. Top and bottom domains at the side-view Figures show TMs close to extracellular (EC) and intracellular (IC) regions, respectively. The number of trajectories used is based on the norm modes available at the NMA (which was around 50 in this study).
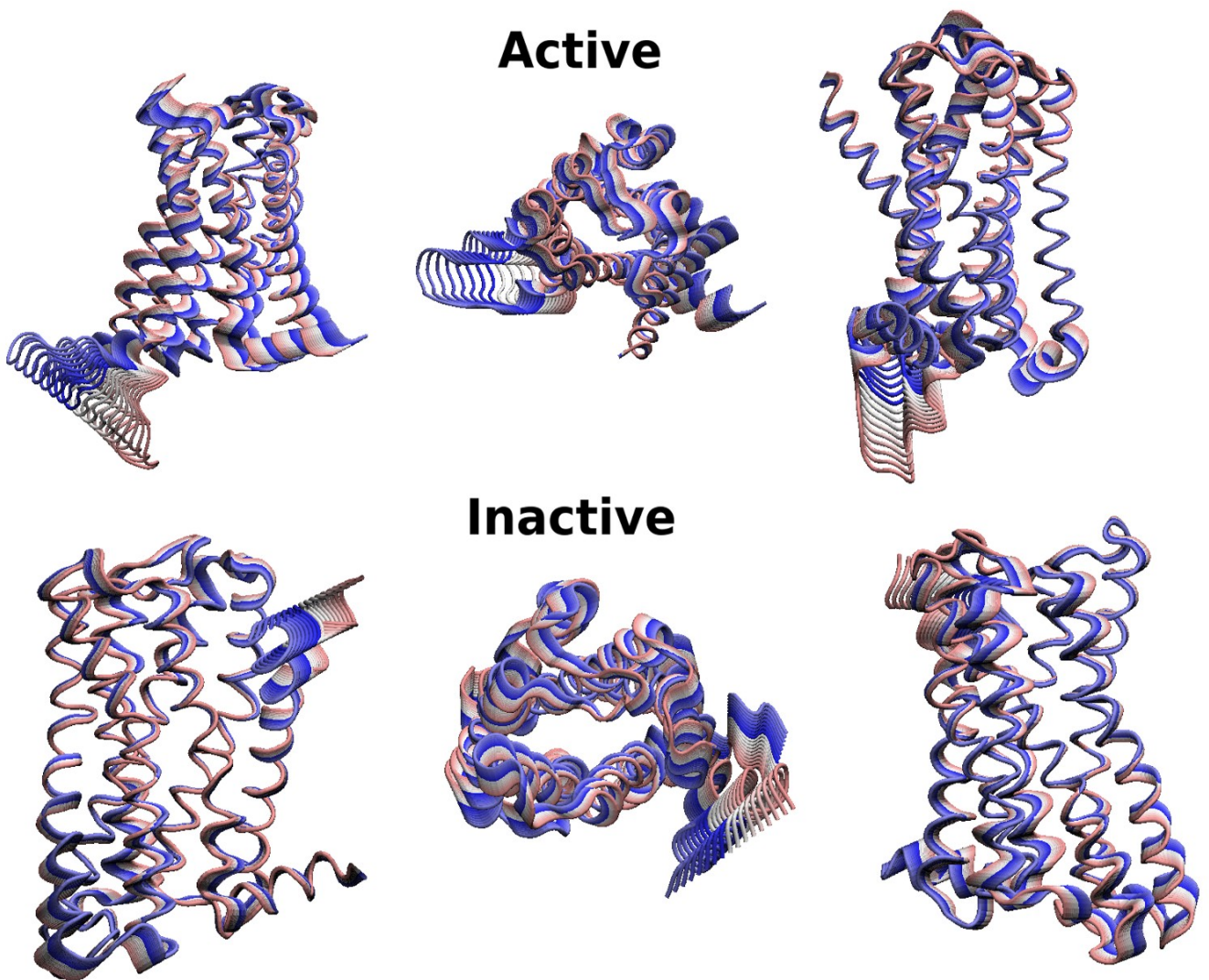
Figure 3. The correlation maps for the D2R receptor models. Black boxes and blank domain represent alpha helices and loop domain, respectively. Colors which are varying from blue (positive correlation) to red (negative correlation) are identifying the intensity of correlation for the each residue.
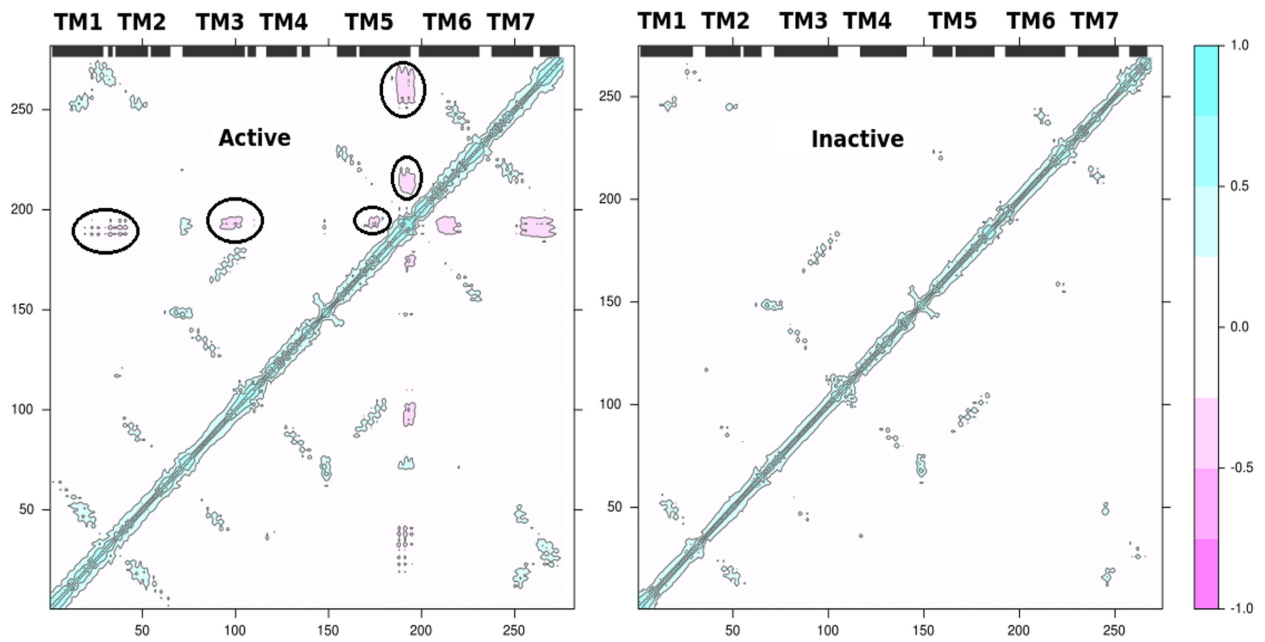
Figure 4. The signal transduction pathway of the active model of D2R based on the NMA. The residues and the domains participated in the negative correlations are displayed as blue balls.
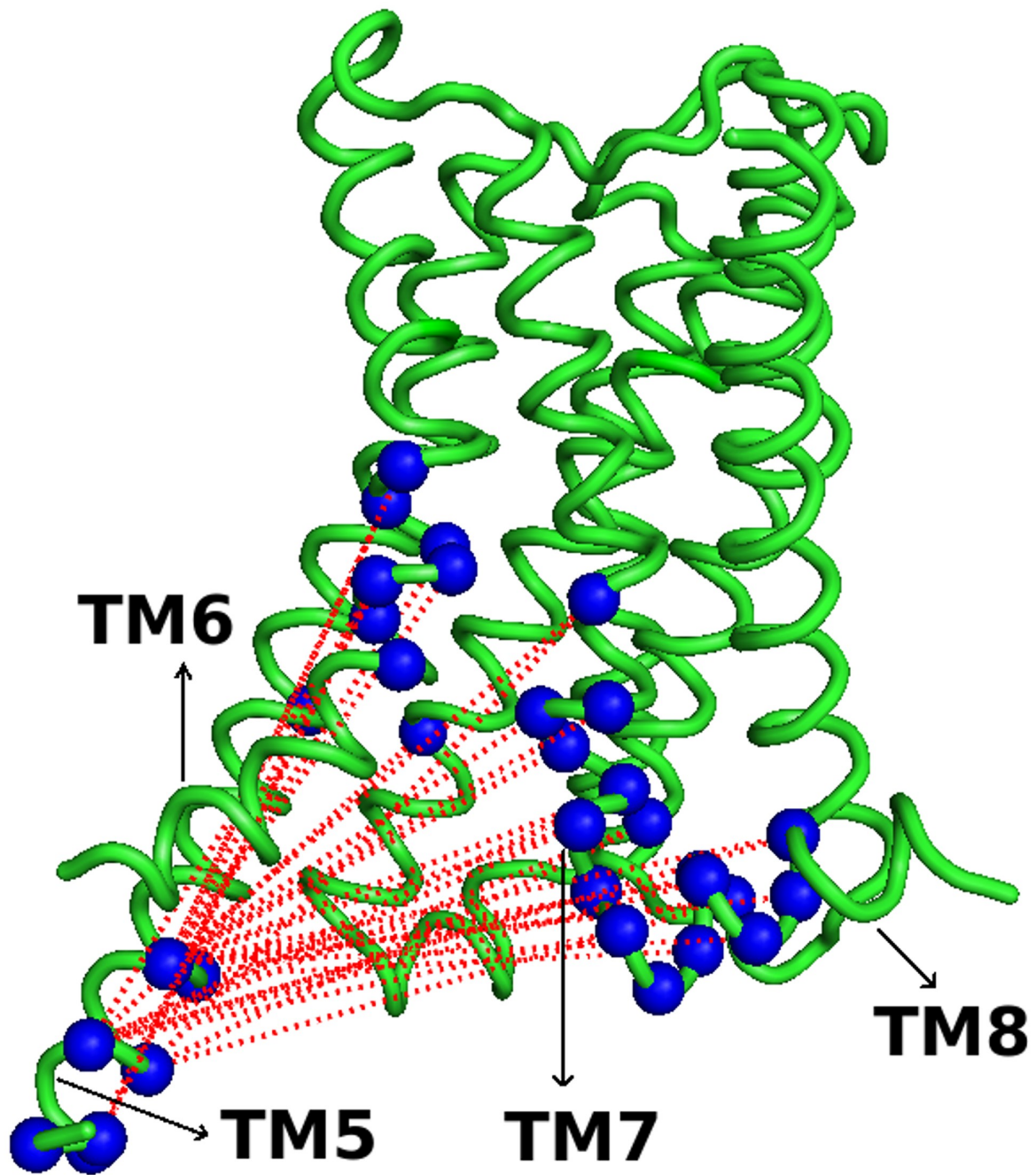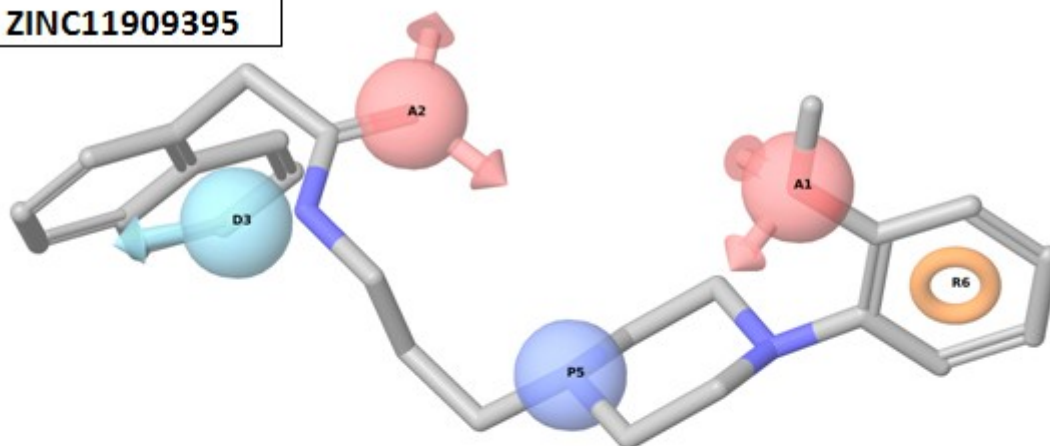
Figure 5. Correlation between the experimental and the predicted activity values (in pIC$_{50}$) of the training set and test set compounds. Labeling of Axes: x-axis experimental, y-axis predicted activity.
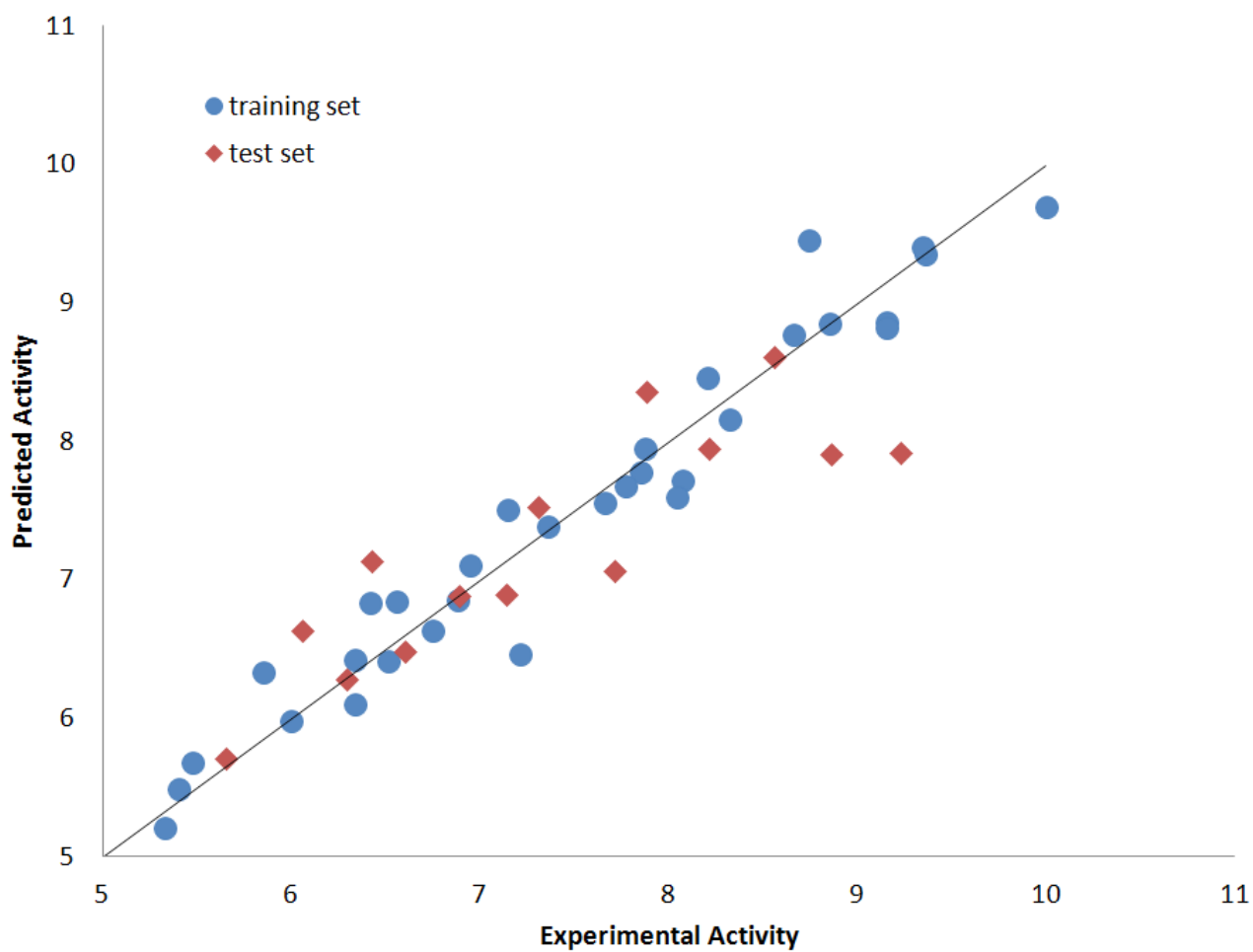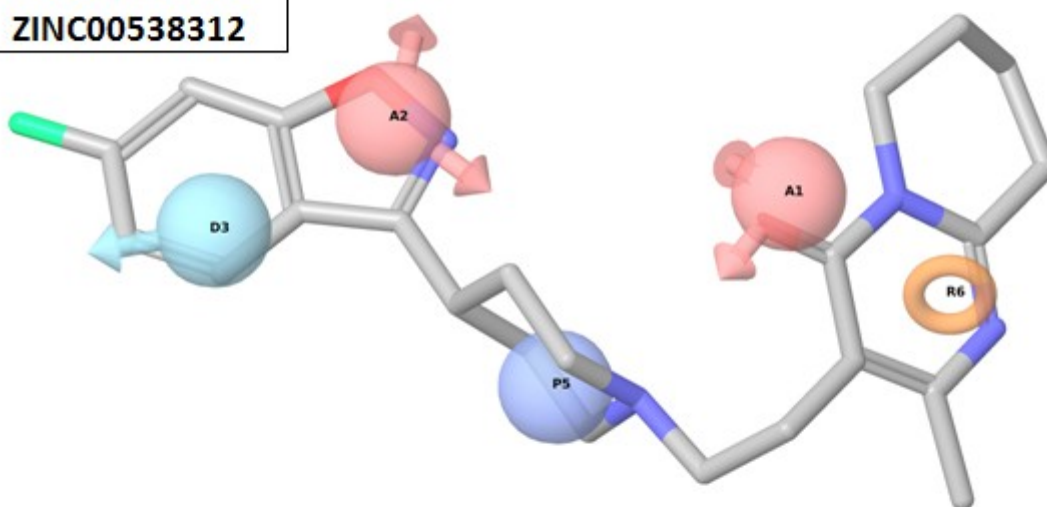
Figure 6. Alignment of the constructed 3D pharmacophore model AADPR.671 with the reference compound, as well as one of selected active (ZINC00538312) and inactive (ZINC10450796) compounds.

ZINC11909395

ZINC00538312

ZINC10450796