UNIVERSITÁ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE DEL FARMACO

SCUOLA DI DOTTORATO IN SCIENZE MOLECOLARI

INDIRIZZO SCIENZE FARMACEUTICHE

CICLO XXVI

TESI DI DOTTORATO

G Protein-Coupled Receptors as a Drug Target: Investigating receptor topology and ligand recognition harnessing the power of Graphical Processing Units

Supervisore: Prof. Stefano Moro.

Direttore della scuola: Prof. Antonino Polimeno.

Dottorando: Davide Sabbadin.

Padova 2011-2014

Ai miei genitori

Abstract

G Protein-Coupled Receptors as a Drug Target: Investigating receptor topology and ligand recognition harnessing the power of Graphical Processing Units

G protein-coupled receptors(GPCRs) constitute a large family of seven domain spanning membrane proteins that mediates a wide variety of cellular processes. Adenosine Receptotors (ARs) are part of this family and are widely distributed through the human body. ARs are involved in the regulation of several physiological processes and their modulation can have potential therapeutic applications for chronic diseases such as Parkinson's and Alzheimer's and for acute conditions such as stroke, cerebral ischemia and cardiac hypoxia. From a computational point of view numerous efforts have been put in place to characterize drug candidates targeting GPCRs. Moreover, the structural information available to the scientific community has assisted to an exponential growth since the determination of the rhodopsin crystal structure. Adrenergic, dopaminergic, histaminergic, opioid and A_{2A} adenosine receptors can provide detailed three-dimensional information useful for supporting structure based drug design approach. We created the first integrated bioinformatics and chemoinformatics web-resource dedicated to Adenosine receptors that is accessible to all the scientific community. It contains an evolutionary driven visualization tool of all Adenosine Receptor models. Adenosiland provides template suggestion in order to get the highest quality receptor model for molecular docking studies and membrane embedded optimized models for biophysical investigation on receptor plasticity. With particular regards to A_{2A} Adenosine Receptor, detailed structural investigation on the dynamic solvation process has been made using state of the art tecnology such as GPU accelerated Molecular Dynamics. Focusing on methodological advances, we report a novel approach consisting in the integration of molecular docking and membrane MD simulations anticipate the bioactive pose of a ligand within the receptor crystallographic structure. Eventually we developed a computational method that enable complete ligand-receptor recognition pathway investigations in a low nanosecond (ns) time scale. We called this new method Supervised Molecular Dynamics (SuMD).
The present research work introduced promising methodological development that can have potential development and implementation on molecular modeling programs that are widely used in both industry and academia.

Sommario

I recettori acccoppiati alle proteine G come potenziali bersagli terapeutici: Investigazione sulla topologia recettoriale e sul riconoscimento ligando-recettore: sfruttando il potere del Processore Grafico

I recettori accoppiati a proteine G costituiscono una grande famiglia di recettori, a sette eliche transmembrana, che media una grande varietà di processi cellulari. I recettori Adenosinici sono parte di questa famiglia e sono distribuiti nella maggior parte dei tessuti del corpo umano. Essi risultatno coinvolti nella regolazione di svariati processi fisiologici. La modulazione dei recettori adenosinici, perciò, può avere potenziali applicazioni terapeutiche per malattie croniche, come il morbo di Parkinson ed Alzheimer, ed acute come infarto, ischemia cerebrale e ipossia cardiaca. Dal punto di vista della chimica computazionale, molti sforzi sono stati compiuti per la caratterizzazione di nuovi candidati farmaci specifici per i recettori accoppiati a proteine G. Inoltre, le informazioni strutturali disponibili hanno assistito ad una crescita esponenziale dalla determinazione della struttura cristallografica della Rodopsina. Recettori adrenergici, dopaminergici, istaminergici, oppioidi e recettori adenosinici, del sottotipo A_{2A}, forniscono informazioni dettagliate per lo sviluppo di approcci di drug-design razionale che sfruttano informazioni riguardanti la struttura molecolare del bersaglio proteico. Abbiamo creato la prima piattaforma web bioinformatica e chemoinformatica integrata dedicata ai recettori adenosinici. Detta piattaforma è a completa disposizione della comunità scientifica e contiene strumenti per la visualizzazione, di tutti i modelli ad oggi clonati, basata su scala evolutiva. Adenosiland fornisce suggerimenti per la selezione del migliore templato, utile alla costruzione di modelli per omologia, allo scopo di compiere studi di docking molecolare. Fornisce inoltre modelli inseriti in un sistema di membrana per investigazioni di natura biofisica sulla plasticità recettoriale. In riferimento al recettore adenosinico A_{2A}, una dettagliata investigazione sul processo di solvatazione dinamico è stata svolta utilizzando studi di dinamica molecolare basati su Processore Grafico (GPU). Inoltre una particolare attenzione è stata posta sull'avanzamento metodologico in chimica computazionale. Riportiamo lo sviluppo di un nuovo approccio che consiste nell'integrazione tra il docking e dinamica molecolare in grado di anticipare la conformazione bioattiva da un vasto insieme di possibili conformazioni di legame nel sito di legame ortosterico del recettore adenosinico umano A_{2A}. Infine è stata sviluppata una nuova metodologia computazionale, chiamata Supervised MD (SuMD), che permette l'investigazione del processo di riconoscimento ligando recettore in una scala dei tempi ridotta, nell'ordine dei nanosecondi (ns). Il lavoro di tesi, qui introdotto, riporta promettenti sviluppi metodologici che possono avere una potenziale implementazione in programmi di modellistica molecolare ampiamente usati in ambiente accademico ed industriale.

Contents

List of abbreviations

ARs: Adenosine Receptors;
DSF: Dynamic Scoring Function;
EL2: Second Extracellular Loop;
EL3: Third Extracellular Loop;
GPCRs: G Protein-Coupled Receptors;
G protein: Guanine nucleotide binding protein;
GPU: Graphical Processing Unit;
hA_1 AR: Human A_1 Adenosine Receptor;
hA_{2A} AR: Human A_{2A} Adenosine Receptor;
hA_{2B} AR: Human A_{2B} Adenosine Receptor;
hA_3 AR: Human A_3 Adenosine Receptor;
IEFs: Interaction Energy Fingerprints;
MD: Molecular Dynamics;
n.d.: not determined;
NECA: N-Ethyl-5'-Carboxamido Adenosine;
POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine;
RMSD: Root Mean Square Deviation;
SAR: Structure Activity Relationship;
SBDD: Structure Based Drug Discovery;
SuMD: Supervised Molecular Dynamics ;
T4E: 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol;
T4G: 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine;
TM: Transmembrane;
WFD maps: Water Fluid Dynamics maps;
ZM 241385: 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-a)(1,3,5)triazin-5-yl-amino)ethyl)phenol.

List of original publications

The thesis report the recent advances in GPU-based Molecular Dynamics simulation applications to Drug Discovery. The presented material is mainly based on the following published articles and submitted manuscripts[1]:

I. Sabbadin, D., Moro, S. *Hydrodynamic 2D/3D-mapping of protein solvation profile using GPU-driven all-atoms Molecular Dynamics* (2014). Journal of Chemical Information and Modeling. Submitted.

II. Sabbadin, D., Moro, *Supervised Molecular Dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale* (2014). Journal of Chemical Information and Modeling. In Press.

III. Sabbadin, D., Ciancetta, A., Moro, S. *Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A_{2A} adenosine receptor as a key study* (2014) Journal of Chemical Information and Modeling. In press.

IV. Floris, M., **Sabbadin, D.**, Ciancetta, A., Medda, R., Cuzzolin, A., Moro, S. *Implementing the Best Template Searching tool into Adenosiland platform* (2014) In Silico Pharmacology. In press.

V. Floris, M., **Sabbadin, D.**, Medda, R., Bulfone, A., Moro, S. *Adenosiland: Walking through adenosine receptors landscape* (2012) European Journal of Medicinal Chemistry, 58, pp. 248-257.

Publications are reproduced with the permission of the authors and are fully reported through the presented work as integrating material. In-depth analysis of developed tools are reported in the appendix section along with the author's full publication list.

[1]Reported articles are chronologically ordered based on the submission date to the scientific journal. Order in which they appear in the thesis may vary. The presented manuscripts contain a Supporting Information section that can be found online.

Preface

G Protein-Coupled Receptors (GPCRs) are transmembrane proteins involved in ubiquitous signal transduction in biological systems. Counting for more than 3% of the genes in the entire human genome [1] those protein machineries link extracellular inputs with diverse cellular responses. This class of proteins regulate crucial cellular functions by responding to a wide variety of structurally diverse ligands, ranging from small molecules (such as biogenic amines, nucleotides, and ions) to lipids, peptides, proteins, and light [2] [3]. It has been estimated that GPCRs constitute the target of about half of the drugs in clinical use today thus, from a structural and pharmacological perspectives, representing an ideal target to design molecules with potential therapeutic effect [4] [5]. Structural biology advances lead to an unpredictable growth resulting in the determination of the high-resolution structures of 16 distinct transmembrane receptors that, including closely related subtype homology models, cover amounts of approximately 12% of the human GPCR superfamily [6]. In particular, Adenosine Receptors, which are part of the family A of GPCRs and comprise four subtypes (A_1,A_{2A},A_{2B} and A_3), regulate key processes of the above-mentioned diseases. At the present time, of all the cloned ARs subtypes, only A_{2A} Adenosine Receptor has been structurally characterized by XRAY spectroscopy. In this such context, expanding GPCRs structural knowledge through Molecular Modeling is a key element to facilitate the drug discovery process. These structural information will help to redefine key concept on GPCRs recognition of such a diverse classes of ligands and elucidate signal transduction across the cell membrane. Moreover, they will provide an enormous opportunity for computational methodologies to make major contributions in this field. MD simulations, harnessing Graphical Processing Units (GPUs) computing power, are closing the gap between theoretical models and experiments thus speeding up the discovery of new chemical entities for the treatment of numerous diseases, including cardiovascular and mental disorders, cancer, and viral infections [7] [8]. In fact, GPU accelerated Molecular Dynamics (MD) simulation based biophysical investigations of biological systems have become a driving factor in molecular pharmacology, improving our understanding of ligand-receptor interaction, activation mechanisms, and receptor hydration [9] [10]. The present thesis covers the recent development of computational approaches that enable the improvement of GPCRs models quality for docking and screening applications, thus facilitating a detailed structural investigation of GPCRs-ligand interaction. Presented methodological advances include investigation of ligand-GPCR recognition process, using Molecular Dynam-

ics simulations, by taking into account the role of water molecules as well as the influence of the membrane on protein flexibility. Along with a brief technical discussion, relevant papers were reported in the appendix section. Several computational tools were applied to study biological systems behavior and protein-ligand interaction at a molecular level. An introductory description of the methods used in this project is presented with particular focus on their application in the drug discovery process.

Homology Modeling

Protein sequences of more than three million proteins are available in the UniProt database [11] [12]. Without further detailing difficulties of structure determinations process for membrane proteins [13], at the end of December 2013 more than 96000 three-dimensional structures of proteins were publicly available in the Protein Data Bank [14]. Considering that a general rule for the folding of a protein has not yet been developed, structural prediction are based on the information available of homologous proteins. The comparative modeling approach, herein referred as Homology Modeling approach, is a computational method based on the notion that the primary structure of proteins is conserved, through evolution, to a lesser extent than the higher-level structures. An amino-acid sequence (target) can be modeled upon the structure of a second protein (template) which are predicted to have the same folding. Based on the sequence alignment of the two proteins, residues are partially matched, taking into account of conserved regions that are evolutionary conserved, and new coordinates are generated.
The generated structural information using the Homology Model procedure can be extremely useful starting point for structure-based ligand/drug design approaches.

Molecular Docking

Molecular Docking is a technique that can enable structural prediction of the binding orientation of small molecules to their protein target by sampling the conformational space within a circumscribed area defined as the binding pocket. The goodness of chemical complementarity between the generated ligand conformations and its receptor is evaluated by an energy function, named scoring function. Clearly molecular docking studies, that are aimed at a better understanding of protein-ligand recognition, are of great interest in medicinal chemistry. The ideal docking protocol would allow both ligand and protein conformational space exploration, with regards of the dynamic solvation process. Despite this, reproducing the conformational space accessible by a macromolecules is computationally demanding. In the normal laboratory practice, ligand conformational space is well sampled while the protein target is treated as a rigid body. Moreover, recent assessments highlighted that solvent treatment is far from far from being realistic and accurate [15] [16].

Molecular Dynamics

Biological Systems, with particular regards to G Protein-Coupled Receptors, are dynamic machineries that respond to a wide variety of structurally diverse ligands. Solvent ionic strength, ligand presence, and membrane environment play an important role in G protein-coupled receptors plasticity. The dynamical behavior of GPCRs is ruled by a balance between intramolecular bonded and non-bonded interactions and the neighboring environment. For example, protein conformational changes in fact can occur after GPCR-ligand recognition [17]. Molecular Dynamics (MD) is a useful simulation technique to investigate physical movements of molecules, during time, based on the numerical integration of the Newton's equations of motion. Deterministic evolution of the conformational change of complex biological systems of relevant pharmaceutical interest have been investigated using this technique [18]. Simulation of large biomolecular systems, however, require the utilization of parallel computers or, recently, commodity Graphical Processing Units (GPUs). The latter represent one of the most important technological breakthrough in computer based simulations allowing to realize the full potential of atomistic simulation without the need of accessing to supercomputers and drastically reducing the cost of science [19]. In fact, Molecular Dynamics simulations on GPU take advantage of the enormous amount of arithmetic units included in each processor and enable a sensible speed up of the simulation as reported in figure 1. Molecular Dynamics simulations, of solvated Dihydrofolate Reductase (DHFR), on a single GPU chip can be as 5 times faster than produced on a single state-of-the-art 16 cores CPU chip and with comparable energy consumption profile.

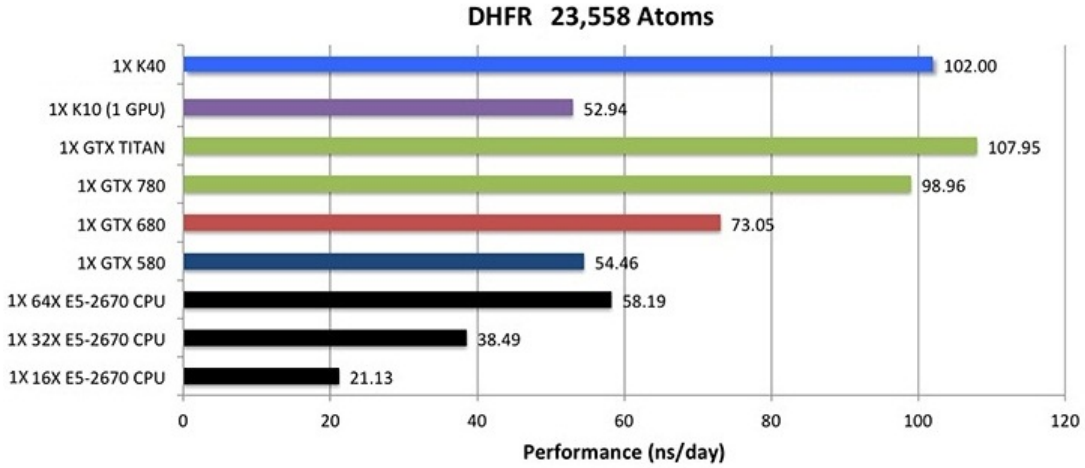


Figure 1: *Molecular Dynamics Single job performances for a single run on a single GPU or CPU chip. The image has been modified with permission from http://ambermd.org .*

With particular regards to the structural information available, membrane protein solved structures in 2010, available in the Protein Data Bank, were 263. In 2014 over 2000 membrane protein solved structures have been published [14]. Conformational dynamics of all protein structures is strictly dependent on the surrounding environment thus investigating the

time-dependent behavior of a molecular system is of great interest. On one hand, structural properties may differ whether a protein is transposed from a crystal to a bilayer environment [20] and the complex liquid crystalline nature of lipid bilayer has proved difficult to map details of protein-membrane interactions using experimental techniques. In contrast, Molecular Dynamics simulations have the potential to provide atomistic-detailed information on protein conformational flexibility, exploration of the ligand conformation within the binding site and characterization of amino-acids interaction with the bilayer. In particular Molecular Dynamics simulation of GPCRs can capture the transient conformational changes between active and inactive state of a GPCR as well as the intrinsic conformational stability of a G Protein-Coupled Receptor model embedded in a lipid bilayer as reported in Figure 2.

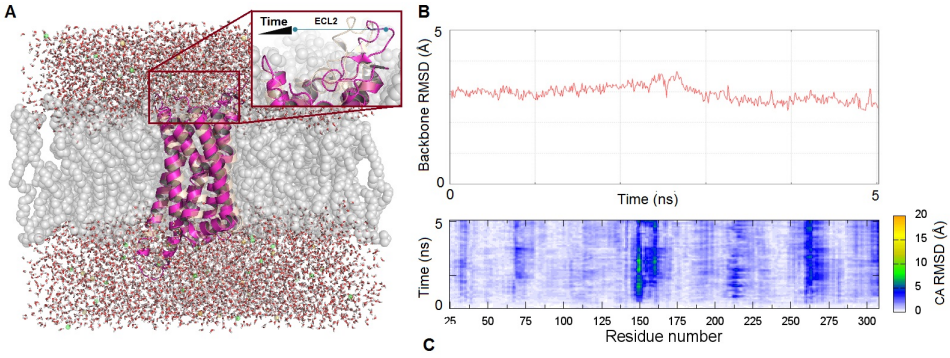


Figure 2: *Panel A. Representation of the dynamic behavior of the rat A_3 adenosine receptor model embedded in a POPC lipid bilayer. Panel B. Backbone R.M.S.D. fluctuation during the unrestrained molecular dynamics simulation. Panel C. per residue CA R.M.S.D. fluctuation during the unrestrained molecular dynamics simulation.*

In this light, membrane MD simulations have the potential to highlight transmembrane domain flexibility related to the overall stability of the protein. This information can be useful to characterize, at an atomistic-level, receptor conformational states during the entire ligand recognition process [17]. Most importantly, the increasing computational power performance and computing architecture diversity allowed Molecular Dynamics simulations to reach timescales comparable with those on which most bio-molecular events of interest take place [18]. In half a decade, since the first microsecond long simulation of Rhodopsin has been reported in literature [21], investigations using all-atom molecular Dynamics simulations have reached the millisecond barrier [8] allowing to investigate the complex recognition process between a ligand and its receptor.

Set up of membrane MD simulation

Molecular Dynamics simulations of membrane embedded receptors, such as GPCRs to perform biophysical studies and to map detailed protein-membrane interaction patterns require precision positioning of the protein, in respect of the lipid bilayer, and precise placement of solvation water molecule and ions. The realization of such complex systems need a multi-step

procedure and comprises the steps herein reported (Figure 3):

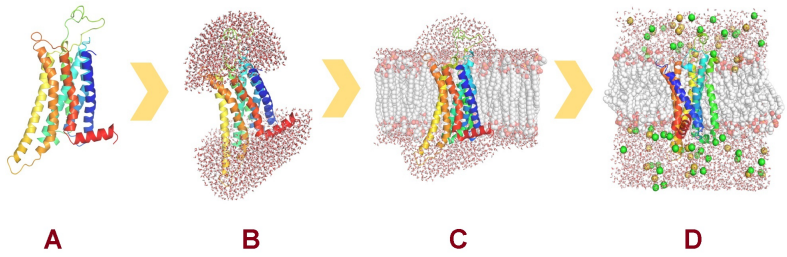


Figure 3: *A typical GPCRs membrane embedding multi-step procedure. The GPCR model is displayed with a ribbon presentation whether explicit lipids, solvent molecules and ions are depicted as spheres and sticks. Details are reported below.*

A. The three-dimensional structure of the transmembrane receptor model is pre-oriented in respect to the normal of the lipid bilayer or using an energy based approach [22];

B. The oriented protein structure is fully solvated with TIP3P water [23];

C. The fully solvated protein is embedded into a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer, according to the suggested orientation, and overlapping lipids and water molecules located into the lipid bilayer (within 0.6 Å) are removed upon insertion of the protein;

D. The system is fully solvated and neutralized by Na^+/Cl^- counter ions to a final concentration of 0.154 M.

After the system is built, in order to calculate the first round of forces applied to the particles, velocities must be assigned. The first velocities assignment are done according to Maxwell-Boltzmann distribution at a give temperature. The system is equilibrated, using a stepwise procedure, in order to reduce steric clashes due to the manual setting up of the membrane-receptor system and to bring the system at a thermic equilibrium. A minimization procedure is suggested. Then, to allow lipids to reach equilibrium and water molecules to diffuse into the protein cavity, the system needs to be equilibrated by keeping the positions of protein and ligand atoms restrained with an harmonic force constant that is gradually reduced and then removed completely. It is strongly suggested to asses the biophysical validity of the built systems by calculating the average area per lipid headgroup (APL) and perform bilayer thickness measurements for each built system. Automated procedures which relies on a similar pipeline can be also performed using CHARMM-GUI membrane builder web service [24].

1

Expanding GPCRs structural knowledge through Molecular Modeling

Contents

1.1 Introduction to G Protein-Coupled Receptors

G Protein-Coupled Receptors possess highly conserved structural features even though the sequence identity among them is rather low. They are composed by seven membrane spanning helices (TM1 to TM7) that are connected by three intracellular (IL1, IL2 and IL3) and three extracellular (EL1, EL2 and EL3) loops. The N-terminal domain is located in the extracellular side whereas the C-terminal domain is located in the cellular cytoplasm. The seven transmembrane helices are the most conserved regions of GPCRs, while N-terminal, C-terminal and loop domains greatly differ in terms of length and function [25] and therefore provide very specific properties to each receptor. The human genome approximately codify for 950 receptors [25], of which 500 are odorant or taste receptors and 450 are sensitive to endogenous ligands [26]. According to sequence analysis GPCRs could be clustered and classified in different classes or families [27] [26] (Figure 1.1). Characterized GPCRs could belong to Family A or rhodopsin-like class, family B or secretin class, family C or metabotropic glutamate and pheromone class, family D or fungal pheromone class, family E or cAMP receptors class or family F or frizzeled/smoothened class [27].

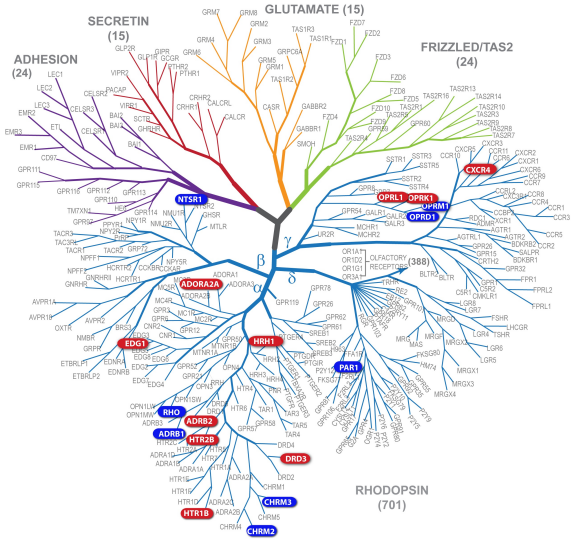


Figure 1.1: *Phylogenetic relationship between the GPCRs in the human genome. Adapted from Katritch et al. [28]*

Among these, family A is the largest and the currently most studied. The overall sequence similarity among all family A receptors is rather low and sequence identity is restricted to a small number of highly conserved key residues (Figure 1.2) located in specific regions in each of the seven transmembrane helices [26] [28]. In details, the extracellular region is responsible for binding diverse ligands and has much higher structural diversity. By contrast, the intracellular region, involved in binding downstream effectors including G proteins and arrestins, is more conserved between GPCRs [28]. Particularly in Adenosine Receptors, which are part of the family A of GPCRs and comprises four subtypes namely A_1,A_{2A},A_{2B} and A_3, highly conserved residues in this class are the DRY motif at the cytoplasmic end of TM3 and two highly conserved cysteine residues in TM3 and in EL2, that form a disulfide bridge [29].

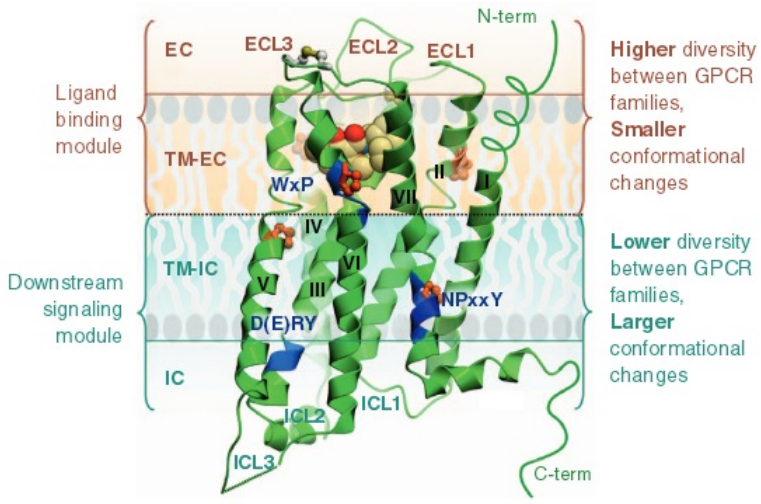


Figure 1.2: *General architecture of GPCRs. Major regions and structural features of GPCRs are shown on an example of the a class A receptor. Blue ribbon patches highlight highly conserved, functionally relevant motifs in the TM helices of class A GPCRs. Adapted from Katritch et al. [28]*

Sharing a common molecular organization, characterized by seven helical trans-membrane domains, serve as a template for the construction of molecular models of other homologous GPCRs 3D models thus representing a powerful tool to better understand the evolutionary path of essential signaling system. Under this perspective, G-Protein Coupled Receptors structural information is the fundamental starting point for all structure-based ligand/drug design approaches.

In the last years, several crystallographic structures of human adenosine A_{2A} receptor in complex with different agonists and antagonists have been solved and released. In particular structural information have been enriched by the physiological agonist adenosine (PDB ID: 2YDO), its N-ethyl-5-carboxamide derivative, NECA, (PDB ID: 2YDV [30]) and the high affinity agonist UK-432097, 6-(2,2-diphenylethylamino)-9-[(2R,3R,4S,5S)-5-(ethylcarbamoyl)- 3,4-dihydroxy-oxolan-2-yl]-N-[2-[(1-pyridin-2-ylpiperidin- 4-yl)carbamoylamino]ethyl]purine-2-carboxamide, (PDB ID: 3QAK [31]) in complex with the human A_{2A} AR. Moreover, the high affinity antagonist (4-(2-[7-amino-2-(2-furyl)[1,2,4]-triazolo[2,3-a][1,3,5]triazin-5-ylamino]-ethyl)phenol, better known as ZM241385, have been co-crystallized with the human Adenosine A_{2A} receptor/T4 lysozyme chimera (PDB ID: 3EML [32]) and with other hA_{2A} AR mutants/chimeras (PDB ID: 3PWH [33] PDB ID: 3VGA [34], PDB ID: 3VG9 [34] and PDB ID: 4EIY [35]). Xantine derivatives such as the N-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl-2,3,6,7-tetrahydro-1H-purin-8-yl)phenoxy] acetamide (PDB ID: 3REY [33]) and caffeine (PDB ID: 3RFM [33]) have been co-crystallized with A_{2A} adenosine receptor. Finally, non-xantine ARs antagonist such as 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (PDB ID: 3UZA [36]) and the 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (PDB ID: 3UZC [36]) have been co-crystallized with the above reported adenosine thermostabilized receptor. Along with the constantly growing of GPCRs structural information, GPCR-dedicated web databases and web services with the aim to have been released to the scientific community [37]. Among them it is surely worth of mention GPCRDB [38], GPCRSSFE database [39], TASSER-907-GPCRs [40], ModeBase [41] and SWISS-MODEL [42] that offer a plethora of pre-generated structural models of GPCRs. Computational web-tools dedicated to G-Protein Coupled Receptors that represent valid scientific resources to build high-quality homology models are GPCR-ModSim [37], MEDELLER [43], CHARM-GUI [44], GPCR-SSFE [39].

1.1.1 Potential therapeutical application of Adenosine Receptors targeting agents

Adenosine Receptors (ARs) are ubiquitously distributed over the human body [45]. In particular the activation of the A_1 AR subtype mediates an inhibition of adenylyl cyclase through activation of pertussis toxin-sensitive $G_{i/o}$ proteins [46] and results in increased activity of PLC [47] [48]. High levels of A_1 Adenosine Receptor expression were found in the brain, heart,

adipose tissue, stomach, testis, spleen, kidney, aorta, liver, eye and bladder [45]Ṙeceptor agonists may find application in various diseases and disorders such as stroke, epilepsy, migraine, pain, cardiac ischemia, arrhythmias, while antagonists could be useful in conditions such as cognitive disorders and edema [45] [49]. Activation of the A_{2A} AR increases adenylyl cyclase activity mediated by G_s. This receptor subtype is also known to act through G_{olf} [50] in the striatum and have shown to activate the PLC pathway in rat artery [51]. The A_{2A} Adenosine Receptors are highly expressed in the striatum, nucleus accumbens, and olfactory tubercle. This receptor subtype has shown to be expressed in immune cells, heart, lung and blood vessels. The therapeutic implications of these receptor subtype agonists result from cardiovascular effects such as vasodilation, tachycardia, hypotension, and platelet aggregation. Additional therapeutic indications for agonists may be respiratory disorders, rheumatoid arthritis, inflammation, wound healing, and sepsis, while antagonists are discussed as treatment in Parkinsons disease, neuronal protection in ischemia, Huntingtons disease and migraine [45]. The A_{2B} Adenosine Receptor is positively coupled to both adenylyl cyclase and PLC [52] [53]. Inhibition of A_{2B} ARs can be useful in diarrhoea, diabetes and asthma. The A_3 ARs have inhibitory effect of adenylyl cyclase [54], stimulation of PLC [55] and calcium mobilization [56] [57]. A protective effect on cardiac cells has shown to be mediated through the activation of K-ATP channels [57]. A_3 adenosine receptor activation may find applications in stroke, asthma, COPD, cardiac ischemia, rheumatoid arthritis and cancer. The blockade of this receptor subtype is useful in glaucoma, asthma and renal failure [45].

1.2 Exploring the Adenosine Receptors landscape

Despite the scientific advance and the rapid growing number of GPCRs structures availability, efforts toward the integration of bioinformatics and chemoinformatics in order to facilitate the exploration of GPCRs from their primary sequences to their three-dimensional structures are required. Moreover, with the growing number of released XRAY structures for the same receptor subtype, a set of rules that can be used in order to select the best structure to use for further structure-based ligand/drug design approaches need to be defined. Eventually, results analysis of the past GPCRdock 2008 competition [58] highlighted that accurate biophysical investigation is also needed to improve GPCRs-ligand models quality for docking and screening applications.

Adenosiland addresses all the presented issues by providing tools for selecting the best template or ARs model to get the highest quality receptor for further molecular docking studies. Moreover, given that phospholipid bilayers are the stage where many essential biophysical and biochemical processes take place, the plasticity of ARs has been investigated using molecular dynamics. Adenosiland can be freely accessed at *http://mms.dsfarm.unipd.it/Adenosiland/*. An in-depth discussion of the most important functionality implemented in Adenosiland can also be found in the Appendix section.

Original publication I

Floris, M., **Sabbadin, D.**, Medda, R., Bulfone, A., Moro, S. *Adenosiland: Walking through adenosine receptors landscape* (2012) European Journal of Medicinal Chemistry, 58, pp. 248-257.

European Journal of Medicinal Chemistry 58 (2012) 248–257

Contents lists available at SciVerse ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Adenosiland: Walking through adenosine receptors landscape

Matteo Floris [a,1], Davide Sabbadin [b,1], Ricardo Medda [a], Alessandro Bulfone [a], Stefano Moro [b,*]

[a] *CRS4, Parco Polaris, 09010 Pula (CA), Italy*
[b] *Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, Via Marzolo 5, I-35131 Padova, Italy*

ARTICLE INFO

Article history:
Received 8 May 2012
Received in revised form
12 October 2012
Accepted 15 October 2012
Available online 23 October 2012

Keywords:
G protein-coupled receptors
Adenosine receptors
Receptor modeling
Bioinformatics platform
Adenosiland

ABSTRACT

Adenosine receptors (ARs) belong to the family of G protein-coupled receptors. Four distinct subtypes are known, termed adenosine A_1, A_{2A}, A_{2B} and A_3. receptors and they are regulated by adenosine which is one of the most ancient and widespread chemical messengers in the animal and plant kingdoms. Moreover, ARs are widely distributed in human body and they are expressed with different density in diverse tissues. It is not surprising that they are involved in the regulation of several physiopathological processes.

Adenosiland represents the first tentative of an integrated bioinformatics and chemoinformatics web-resource dedicated to adenosine receptors. This informatics platform provides a wide-ranging of structure based and ligand based query functions to facilitate the exploration of adenosine receptor structures from primary sequences to three-dimensional architectures. Here, we present an overview of *Adenosiland* platform describing the most valuable searching tools and their functionalities. *Adenosiland* can be freely accessed at *http://mms.dsfarm.unipd.it/Adenosiland/*.

1. Introduction

Purines (most notably ATP and adenosine) appear to be the most ancient and widespread chemical messengers in animal and plant kingdoms [1]. This different purinergic transmitters act upon target cells through activation of three classes of receptors: the metabotropic P1 receptors to adenosine, and nucleotide receptors of the P2 family, which is further subdivided into P2Y metabotropic and P2X ionotropic sub-classes [2]. In particular, the P1 class comprises four types of G protein-coupled adenosine receptors A_1, A_{2A}, A_{2B} and A_3 [3]. These receptors are generally coupled to adenylate cyclase. Activation of the A_1, and A_3 receptors has an inhibitory effect, whereas A_{2A} and A_{2B} stimulate production of cyclic AMP (cAMP) [3]. As anticipated, the purinergic signaling can be considered ubiquitous. Although our knowledge of chemical sensitivity of bacteria is quite fragmentary, there is significant evidence demonstrating that purines and pyrimidines exert a wide range of actions on bacteria. For example, adenosine inhibits growth of several bacteria species [4]. Following the evolutionary path, there are several similarities between insect and mammalian adenosine receptor functions: extracellular adenosine influences immune responses in both; adenosine agonists and antagonists modulate the sleep and waking cycle in Drosophila, perhaps associated with the endogenous expression of adenosine receptors in the insect brain [5]. Moreover, adenosine, plays a central role in this process in most advanced snakes, birds and all mammalians [1–4]. In human beings, the autacoid adenosine plays a pivotal role in a large variety of physiological and physiopathological processes both in central nervous system (CNS) and in periphery [6]. As already described, adenosine is physiologically present in the extracellular fluid and exerts its effects through activation of four cell surface receptor subtypes which belong to the superfamily of G protein-coupled receptors. Adenosine receptors (ARs) are widely distributed in the body and are expressed with different density in diverse tissues [7]. The classical transduction intracellular pathways associated with AR stimulation are inhibition, via $G_{i/o}$ protein (A_1, and A_3. subtypes) or activation, via G_s protein (A_{2A} and A_{2B} receptors), of adenylate cyclase (AC) [7]. More recently, other second messenger systems, such as phospholipase C or potassium and calcium channels, have been described as relevant for AR signaling. As all other members of the GPCR family, ARs share a similar structural architecture consisting of seven trans-membrane helices that contain well-conserved sequence motifs [7].

In the last few years, several crystallographic structures of human adenosine A_{2A} receptor in complex with different agonists and antagonists have been solved and released. The ARs physiological

* Corresponding author. Tel.: +39 049 8275704; fax: +39 049 8275366.
E-mail address: stefano.moro@unipd.it (S. Moro).
[1] These authors equally contributed to the work.

0223-5234/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.ejmech.2012.10.022

agonist adenosine (PDB ID: 2YDO), its *N*-ethyl-5′carboxamide derivative, NECA, (PDB ID: 2YDV [8]) and the high affinity agonist UK-432097, 6-(2,2-diphenylethylamino)-9-[(2R,3R,4S,5S)-5-(ethylcarbamoyl)-3,4-dihydroxy-oxolan-2-yl]-*N*-[2-[(1-pyridin-2-ylpiperidin-4-yl)carbamoylamino]ethyl]purine-2-carboxamide, (PDB ID: 3QAK [9]) have been co-crystallized with the human A_{2A} AR. Moreover, antagonists belonging to different chemical families, have been also co-crystallized with the human A_{2A} AR. In particular; the high affinity antagonist (4-(2-[7-amino-2-(2-furyl) [1,2,4]-triazolo[2,3-α] [1,3,5]triazin-5-ylamino]ethyl)phenol, better known as ZM241385, is in complex with the human Adenosine A_{2A} receptor/T4 lysozyme chimera (PDB ID: 3EML [10]) and in complex with other hA_{2A} AR mutants/chimeras (PDB ID: 3PWH [11] PDB ID: 3VGA [12], PDB ID: 3VG9 [12] and PDB ID: 4EIY [13]).

Finally, xantine derivatives such as the *N*-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl-2,3,6,7-tetrahydro-1H-purin-8-yl)phenoxy]acetamide (PDB ID: 3REY [11]) and the very well known caffeine (PDB ID: 3RFM [11]) have been co-crystallized with A2A adenosine receptor. Recently, ARs structural information have been furthermore enriched by the co-crystallization of 1,2,4-triazine derivatives such as the 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (PDB ID: 3UZA [14]) and the 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (PDB ID: 3UZC [14]) with a thermostabilised human adenosine A_{2A} receptor.

From biophysics and pharmacological perspectives, GPCRs structural information represent a powerful tool to better understand the evolutionary path of this ancient and essential signaling system. Moreover, X-ray information can be considered the fundamental starting point for all structure-based ligand/drug design approaches [15]. In fact, sharing a common molecular organization characterized by seven helical trans-membrane domains any crystallographic structure can in principle serve as a template for the construction of molecular models of other homologous GPCRs.

With the development of computer technology and its use by biologists, pharmacologists and medicinal chemists, bioinformatics and chemoinformatics gradually become powerful disciplines to correlate protein structures with protein functions or dysfunctions. In this context, *Adenosiland* represents the first tentative of an integrated bioinformatics and chemoinformatics web-resource dedicated to adenosine receptors. One of the most interesting added value of *Adenosiland* is that its informatics infrastructure can be easily adaptable to other GPCR families. In fact, this platform provides a wide-ranging of query functions to facilitate the exploration of GPCRs from their primary sequences to their three-dimensional structures.

Even if several GPCR-dedicated web servers and web services are already available (as described below), we would like to underline that *Adenosiland* is not only just "another" GPCR visualization tool but it represents the assembling of the state of the art of different bioinformatics, chemoinformatics and molecular modeling approaches into a single virtual chamber. In *Adenosiland*, any scientist can easily compare adenosine receptors from an evolutionary point of view or, in alternative, from a functional point of view. In fact, starting from the available crystallographic information, all known adenosine receptor structures have been derived by using homology modeling technologies, and they have been further optimized in their native membrane environment using molecular dynamics simulations. Receptor models of cloned adenosine receptors can be intuitively visually inspected, analyzed in real time and downloaded.

Finally, we have implemented an useful tool to provide template suggestions and homology models of all four human adenosine receptors based on the "similarity" between an external agonist/antagonist and all co-crystallized adenosine ligands. We consider this information a crucial starting point for further molecular docking studies. In fact, the selection of the best template to build appropriate homology models, incorporating the ligand induced-fit on the receptor counterpart, is a key feature to facilitate the identification or the optimization of novel potent and selective agonists and antagonists.

Here, we present an overview of *Adenosiland* platform describing the most valuable searching tools and their functionalities. *Adenosiland* can be freely accessed at *http://mms.dsfarm.unipd.it/Adenosiland/*.

2. Materials and methods

2.1. Implementation

Adenosiland infrastructure, based on Ubuntu 9.10 Linux operating system, is a patchwork of several informatics tools including: Apache web server version 2.0 [16]; PHP scripting language [17]; Jmol, the open source molecular viewer [18]; Jquery, Javascript library [19]; Jalview, a Java Alignment Editor [20,21]; Yasara modeling suite (version 11.4.18) [22]; Molecular Operating Environment suite (MOE, ver. 2010.10) [23]; MEGA 5.0 [24] and RdDe, the residue based diagram editor [25]. In addition CDK, the chemistry development kit [26,27]; CACTVS toolkit [28]; Indigo command line utilities [29]; Align-it [30] and pepMMsMIMIC scripts [31] have been incorporated in the core of *Adenosiland* architecture. We desire to give our appreciations to Peter Ertl for his courtesy in using its the Java Molecule Editor (JME).

2.2. Homology modeling

Two different homology modeling protocols have been used for the construction of adenosine receptor models.

- Yasara Structure Suite [22] has been used to build receptor models, collected in both *AdeList* and *EvoTree* sections of *Adenosiland*, and for their membrane embedding and structural comparison. The crystal structure of human adenosine A_{2A} receptor bound to the high affinity antagonist ZM241385 (PDB ID: 3EML) [10] has been used as template structure. The lysozyme portion fused to the receptor has been removed before starting the homology model procedure. FASTA sequences of all cloned adenosine receptors were retrieved from UniProtKB/Swiss-Prot [32–34] and analyzed using PSI-BLAST [35], PSI-PRED and SSALN [36] to establish the most accurate alignment against the template. N-terminal and C-terminal were deleted if their lengths exceed those found in the crystallographic template.

  Particular attention has been dedicated to the reconstruction of both extracellular loop 2 (ECL2) and intracellular loop 3 (IL3). They were extensively sampled using the loop search tool implemented in Yasara suite [37]. Side chains of the entire model were optimized using SCWALL method [38] in tandem with YASARA2 force field [39]. All receptor models were energetically optimized until the convergence criteria of 0.05 kJ/mol per atom was reached. Disulfide bridges have been incorporated in the homology models following both crystallographic and mutagenesis information. In particular, according to Schiedel and collaborators [40], only one cysteine bridge, links Cys78 (3.25) to Cys171 (ECL2) in all A_{2B} receptor models.
- All human A_1, A_{2A}, A_{2B} and A_3 adenosine receptor models, downloadable from the "*Best Template Searching*" section of *Adenosiland*, were constructed using as template structures all the published crystallographic structures of human adenosine A_{2A} receptors available at the time of the preparation of this

manuscript (in total 12 structures: PDB ID: 2YDO, 2YDV [8], 3EML [10], 3QAK [9], 3PWH, 3REY, 3RFM [11], 3UZA, 3UZC [14], 3VGA, 3VG9 [12], 4EIY [13]). The eventual lysozyme/antibody portion fused to the receptor, co-crystallized ligands and water molecules have been deleted before starting the homology model procedure. "Protonate-3D" tool was used to appropriately assign ionization states and hydrogen positions [41]. To minimize contacts among hydrogens, the structures were subjected to Amber99 [42] force field minimization until the *r.m.s.* of conjugate-gradient was <0.05 kcal mol^{-1} Å^{-1}, keeping the heavy atoms fixed at their crystallographic positions. FASTA sequences were aligned, using Blosum50 matrix, with the template sequence. Backbone and conserved residues coordinates were copied from the template structure therefore newly modeled regions and non conserved residues side chains were modeled and energetically optimized, using Amber99 force field [42], until the *r.m.s.* of conjugate-gradient was <0.05 kcal mol^{-1} Å^{-1} was reached. Disulfide bridges have been incorporated in the homology models as described above. Missing loop domains were constructed by the loop search method implemented in Molecular Operating Environment (MOE, version 2010.10) program [23]. N-terminal and C-terminal were deleted if their lengths exceed those found in the crystallographic template. "Protonate-3D" tool was used to assign ionization states and hydrogen positions [41]. Protein stereochemistry evaluation was then performed by several tools (Ramachandran and χ plots measure j/ψ and $\chi1/\chi2$ angles, clash contacts reports) implemented in MOE suite.

Molecular graphics were created with YASARA [22] and POV-Ray [43].

2.3. Membrane molecular dynamics

All modeled receptors were embedded in a membrane environment using a palmitoyl-oleoyl-phosphatidyl-choline (POPC) bilayer. The orientation of each receptor in membrane is obtained from the "Orientations of Proteins in Membranes (OPM)" database [44]. The solvent exposed area has been solvated with TIP3P water [45] using the program Solvate 1.0 [46]. Overlapping lipids (within 0.6 Å from protein) and eventual water located in the hydrophobic protein-membrane interface (within 3 Å from lipids molecules) were removed upon insertion of the protein. The final complex has been electrically neutralized with a total ionic concentration (Na^+ and Cl^- ions) of 0.154 M.

Molecular dynamics simulations were carried out using ACEMD program [47] on a local GPU cluster. An harmonic restraint 1 kcal mol^{-1} Å^{-2} was applied to the backbone atoms of the receptor during the first 1000 steps of conjugate-gradient minimization. This harmonic restraint has been also maintained during the first 10 ns of dynamics simulation, then scaled to 0.1 kcal mol^{-1} Å^{-2} during the further 15 ns. Temperature was maintained at 310 K using a Langevin thermostat with a low damping constant of 1 ps^{-1}. Pressure was maintained at 1 atm using a Berendensen barostat. The system was finally equilibrated till 20 ns in the isothermal-isobaric ensemble (NPT) without applying restraints to the atom of the built system, under periodic boundary conditions. The long-range Coulomb interaction was handled using the particle mesh Ewald summation method (PME) [48] with grid size rounded to the approximate integer value of cell wall dimensions. A non-bonded cutoff distance of 9 Å was used with a switching distance of 7.5 Å. For the equilibration runs, the M-SHAKE algorithm [49] has been used on all atoms covalently bonded to a hydrogen atom with an integration time step of 2 fs. All molecular dynamics simulations were performed using Charmm27 [50] force field. The dimensions of the simulation box resulting from NPT equilibration procedure are specified in the CRYST parameter inside every downloadable PDB file. Ramachandran plots for the built models were obtained using RAPPER service [51].

3. Results and discussion

3.1. General features of Adenosiland components and tools

As anticipated, the main focus of *Adenosiland* is to create a virtual space where simultaneously analyze sequence and structural information assigned to all cloned adenosine receptors.

Adenosiland is directly connected to UniProt database [32] where it is possible to automatically download all unique cloned sequences of all four adenosine receptors subtypes. At the time of the preparation of this manuscript, *Adenosiland* collects 29 unique sequences clustered in nine adenosine A_1 receptors, six A_{2A} receptors, seven A_{2B} receptors, and seven A_3. As soon as a new adenosine receptor sequence becomes available from UniProt database, *Adenosiland* is immediately alerted and updated.

As previously mentioned, despite the recent and successful efforts in crystallization of GPCR proteins, homology modeling approach is still widely used as a method for obtaining preliminary structural information of other receptors. With the aim of bridging the gap between the number of sequence and the number of three-dimensional structures, a robust homology modeling approach has been used to populate *Adenosiland* with the corresponding receptor models, as described in detail into Materials and Methods paragraph. Indeed, evaluation and validation of homology models are indispensable in particular for membrane proteins such as GPCRs. Molecular dynamics (MD) simulations in a lipid bilayer environment provide a possible tool to address the latter aspect. Following this approach, all adenosine receptor models have been embedded in a membrane environment and equilibrated for 20 ns. An example of MD trajectory analysis has been reported in the Supplementary Information section (see Fig. S1). We have not deeply analyzed our MD simulations only because we consider 5 ns not enough to extrapolate robust information from the corresponding trajectories. Indeed, we consider our embedded receptor models as an alternative structural starting point to the more conventional homology models. Users can freely download all membrane receptor models and use them for any further type of biological or biophysical studies.

3.2. Why Adenosiland

Along with the constantly growing of GPCRs structural information, an increasing number of GPCR-dedicated web databases and web services have been released to the scientific community.

Among them it is surely worth of mention GPCRDB [52], GPCR-SSFE database [53], TASSER-907-GPCRs [54], ModeBase [55] and SWISS-MODEL [56] that offer a plethora of pre-generated structural models of GPCRs. Other useful computational web-tools dedicated to GPCRs are GPCR-ModSim [57], MEDELLER [58], CHARM-GUI [59], GPCR-SSFE [53]. An exhaustive summary on this topic has recently been covered by Rodriguez and collaborators [57].

As already mentioned in the Introduction, *Adenosiland* is not a simple depository of receptor models or an alternative visualization web tool. We desire to consider *Adenosiland* a virtual space where bioinformatics bridges chemoinformatics in the context of adenosine receptors. Interestingly, the informatics infrastructure behind *Adenosiland* can be easily adaptable to other GPCR families. The platform represents a starting point for non experienced modeler users that are interested in exploring adenosine receptor architectures, walking across the evolutionary pathway of this

specific metabotropic receptor family, or analyzing the structural differences among the four diverse receptor subtypes.

However, *Adenosiland* is also meant to be a starting point for more experienced users. The possibility to download all human adenosine receptor homology models, built accounting for induced-fit by the ligand on the receptor, offers a ready-to-run starting point for further receptor-driven ligand design studies for the identification or optimization of novel potent and selective agonists and antagonists of adenosine receptors.

Moreover, from a biophysics point of view users have the possibility to download all necessary files to perform molecular dynamics simulations of any modeled adenosine receptor, embedded and equilibrated in a lipid bilayer, allowing the explore the time-dependent conformational behavior of the receptor at atomistic level.

3.3. *Adenosiland: database organization*

A schematic flow chart of *Adenosiland* architecture is shown in Fig. 1. All sequences and structural information are stored and organized inside the *Adenosiland* database which it is easily accessible from the main web page by three alternative search process tools called *AdeList*, *EvoTree* and *Best Template Searching*, as shown in Fig. 2. Specifically, *AdeList* represents a conventional receptor subtypes table in which all processed receptor subtypes are listed based on their receptor membership. On the bottom part of *AdeList* webpage, there is the option to quickly download all three-dimensional structures in PDB format of all receptors subtypes in vacuum or in membrane conditions. Moreover, *EvoTree* is an alternative evolutionary-driven selection tool where all receptors are graphically grouped and ordered according to their evolutionary distances, so all receptors with a distance of less than 0.02 to the group's common ancestor were grouped together. In *EvoTree*, it is possible to rapidly select all receptors which belong to a specific evolutionary branch by clicking to the corresponding evolutionary node.

Any receptor selection obtained using both *AdeList* and *EvoTree* tools is processed and all sequences and structural information are displayed in a new webpage where it is possible to compare the primary sequence alignment of the preselected receptors and also the

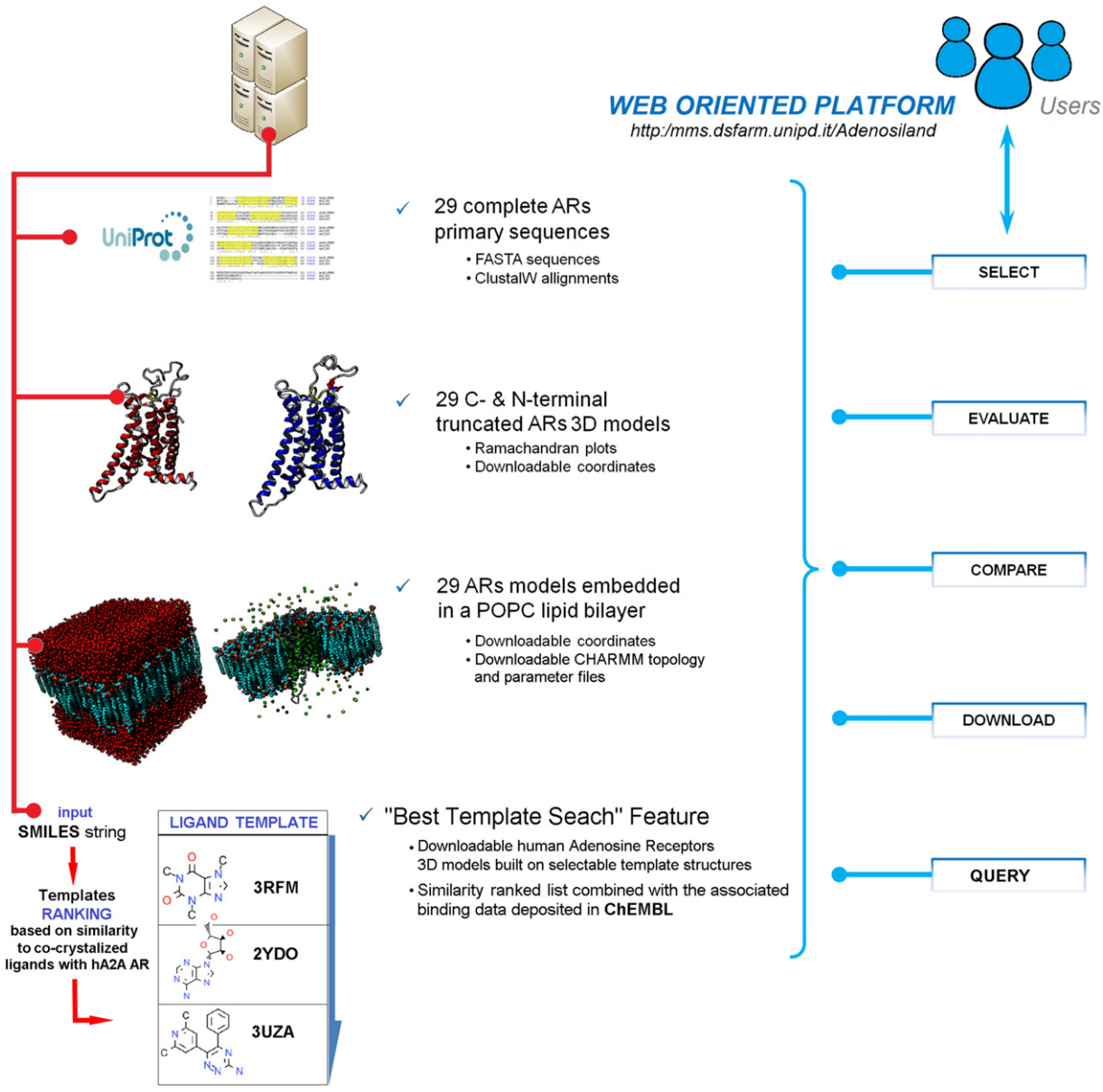


Fig. 1. Schematic flow chart of *Adenosiland* architecture.

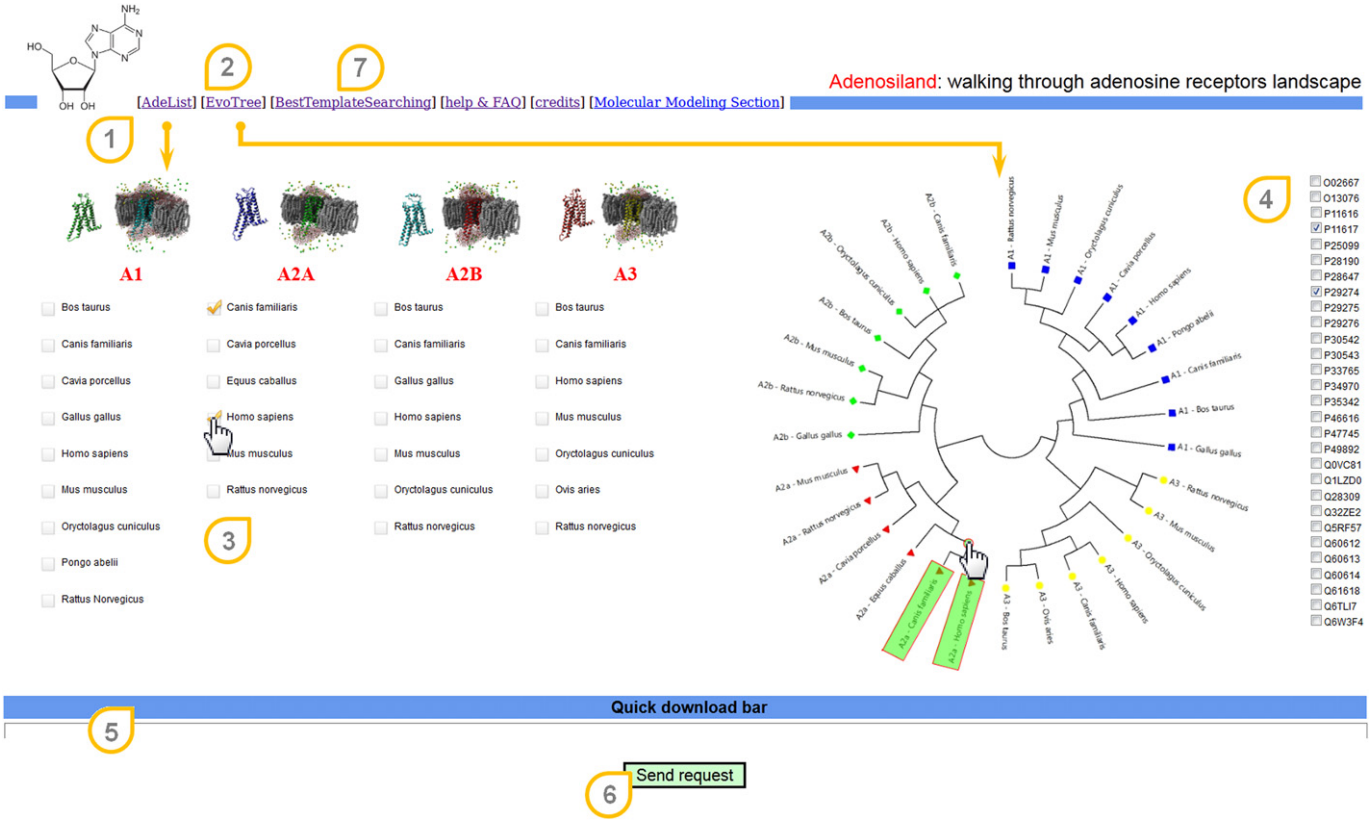


Fig. 2. *Adenosiland* web interface composed by three alternative search process tools: *AdeList* (1), *EvoTree* view (2) and *Best Template Searching* tools (7). In *AdeList* (3) all modeled receptor subtypes are listed based on their subtypes receptorship. *EvoTree* is an alternative evolutionary-driven selection tool (4). *Best Template Searching* (7) provides template suggestions and homology models of all four human adenosine receptors, based on similarity between an external agonist/antagonist and all co-crystallized adenosine ligands. A practical download bar is located at the bottom of *AdeList* page (5). The "Send Request" button (6) gives access to an interactive page where information about selected entries are available.

superposition of their homology modeling-driven three-dimensional structures, as shown in Fig. 3. Interestingly, it is possible to interactively move from the sequence-based amino acids selection to the corresponding visualization into the three-dimensional view. A number of pre-cooked selection/colouring tools are available to facilitate a better browsing experience and access to more valuable receptor information. Users can specifically locate differences in the aligned primary sequences, between the selected set of adenosine receptors, and have an instant visual feedback about the three-dimensional localization of such differences using the built in interactive viewer. Moreover, the most important bioinformatics information is also shown into the right side of this specific report page.

Finally, we have implemented an novel tool, called "*Best Template Searching*" to provide template suggestions and homology models of all four human adenosine receptors based on the "similarity" between an agonist or antagonist, provided by the user, and all co-crystallized adenosine ligands. In fact, in most known cases, ligand topology shapes its target protein upon binding depending on its size/shape and on the interaction network with the amino acids of its orthosteric site. In fact, the ligand-driven induced-fit of the receptor is a key feature to facilitate the identification or the optimization of novel potent and selective agonists and antagonists, in particular through molecular docking studies.

Using the "*Best Template Searching*" option, users can upload a SMILES string or directly draw the 2D structure using JME interface of its favorite scaffold and search, by similarity, the closest ligand already co-crystallized with the human A2A receptor. Several similarity indexes were calculated using different approaches such as a 2D similarity, calculated from Pubchem Fingerprints (CDK implementation), based on Tanimoto and Tversky indexes [26,27], a shape similarity calculated by using an in-house implementation of the Ultrafast Shape Recognition method [31,60], a pharmacophore-based similarity, where pharmacophoric features are described by Gaussian 3D volumes [30], and simple consensus shape- and pharmacophore-based similarity index derived by the following function: $0.6 \cdot \text{pharmacophoric similarity} + 0.4 \cdot \text{shape similarity}$. The values of the two coefficients have been obtained by a preliminary in-house validation based on all available crystallographic structures (data not shown). Based on the chosen similarity sorting, all human adenosine receptor subtypes can be downloaded and used for further receptor-based ligand design strategies such as for docking-driven virtual screening or pharmacophore searching.

Simultaneously to the best template searching process, a similarity search screening is also performed against all adenosine agonists and antagonists deposited in ChEMBL [61] Similarity search finds known adenosine ligands with a high percentage of features that are common to the target molecule using several similarity metrics as described in details into Material and methods section. A similarity ranked list combined with the associated binding data available in literature are organized in a table as shown in Fig. 4.

3.4. Bioinformatics and structural information deliverable by Adenosiland

Considering the amount of data collected into *Adenosiland*, in this paragraph we would like to summarize the most relevant information deliverable from a simple navigation through the

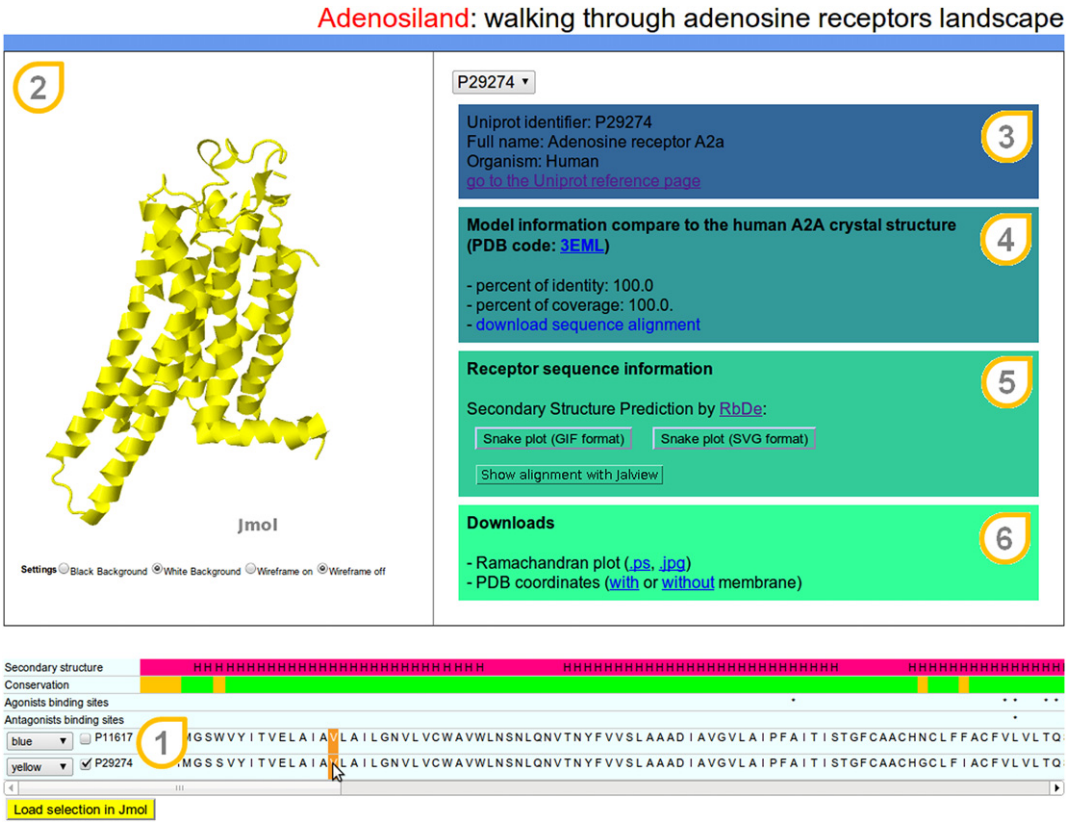


Fig. 3. *Adenosiland* interactive web interface. Primary sequence alignment of the preselected receptors and superposition of their homology modeling-driven three-dimensional structures (1). Embedded Jmol visualizer (2) allows to offer various visualization possibilities. Enhanced browsing experience allows access to more valuable receptor information (3, 4 and 5). The download button (6) gives the opportunity to access and inspect data to the scientific community.

adenosine receptors biological space. Some of the most relevant bioinformatics information are collected in Table 1. Starting from the simple primary sequence analysis, among adenosine receptors the A_{2A} subtype is characterized by a longer primary sequences mainly due to its C-terminal tail (>110 amino acids). It is very well documented that the C-terminus of adenosine A_{2A} receptor is a pretty crowded place considering the number (at least five) of accessory proteins that interact with, and this role seems to be conserved in all the analyzed adenosine A_{2A} receptors considering the very high similarity of their C-term domains [62]. Another crucial architectural element among all GPCRs is the second extracellular loop (ECL2) that may orchestrate a network of interactions that may stabilize the inactive conformation of the receptor and/or kinetically control the kon/koff ratio of the receptor–ligand recognition [63]. In our adenosine receptors ensemble, A_{2B} subtypes are characterized by the longest ECL2 (≥38 amino acids) where, in A_3 subtype, ECL2 is the shortest (≥28 amino acids).

In this specific context, despite the high degree of structural diversity with respect to ECL2 in family A GPCRs, there is one feature that is conserved in the vast majority of GPCRs a disulfide bond between ECL2 and the top of TM3 (Cys3.25). This disulfide bond effectively tethers ECL2 on the top of the TM helical bundle and provides a very important conformational constraint of the ECL2. Some GPCRs have additional disulfide bonds between different ECLs such as for example between ECL2-ECL1 in A_{2A} subtype. The two cysteines involved in the formation of this additional ECL2-ECL1 disulfide bridge are conserved in all analyzed A_{2A} receptors. Additionally, the A_{2A} subtype also possess an additional intra-loop disulfide bond within ECL3, in common with melanocortin receptors and human histamine receptor 1. These "additional" disulfide bonds contribute to reduce the flexibility of ECLs and, consequently, they peculiarly sculpt the topography of the extracellular portion of the receptor in proximity of the orthosteric binding cleft. Finally, according to Schiedel and collaborators, only one cysteine bridge, linking TM3 to ECL2 in A_{2B} receptor models, is detectable [40]. This information has been taken into consideration during the homology model building procedure of all A_{2B} receptors. A summary of conserved cysteines putatively involve in disulfide bridges is listed in Table 2. All modeled receptors have been constructed following disulfide bridges information reported in Table 2.

The extracellular face of GPCRs is usually *N*-glycosylated, with at least one glycosylation site (N-X-S/T) on the N-terminus, often more. Moreover, it is known that *N*-glycosylation of ECL2 is quite common in GPCRs: for example, A_3 receptors from sheep (*Ovis aries* P35342), dog (*Canis familiaris* Q28309) and mouse (*Mus musculus* Q61618) share a potential glycosylation sites on ECL2. Besides, A_{2A} and A_{2B}, A_1 subtypes lack of N-terminal glycosylation sites but maintain the glycosylation site at the ECL2.

In Table S1 (see Supplementary Information), we have summarized the degree of conservation of the most crucial amino acids involved in the recognition of both agonists and antagonists. It is worth to underline that among all adenosine receptor subtypes the residues of the putative agonist/antagonist binding sites are largely conserved. Comparing all receptor–ligand contacts, it is curious that Trp6.48, located at the bottom portion of TM6 of the orthosteric pocket and thought to have a crucial involvement in activation and deactivation receptor mechanisms, is replaced by a cysteine residue in the A_{2A} receptor of *guinea pig* [10,64]. Moreover His6.52, also implicated in the activation process of the receptor, is mutated into a serine residue in all A_3 receptor subtypes and in the A_{2A} receptor of *guinea pig* [65]. Beside these two

2	3	3		1		4						
CHEMBL COMPOUND ID	ADENOSINE REC	CHEMBLTARGET ID	ORGANISM	IC50 uM	SMILES	USR	PUBCHEM TANI	PUBCHEM TV	PHARAO TANI	PHARAO TV	COMBINED	CHEMBL ASSAY ID
CHEMBL464859 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.00075	CCNC(=O)[C@H]1O....	0.46	0.80	0.91	0.18	0.71	0.61	CHEMBL636672
CHEMBL68738 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.001	OC[C@H]1O[C@H]....	0.30	0.81	0.91	0.17	0.70	0.54	CHEMBL636672
CHEMBL85549 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.006	OC[C@H]1OC([C@H](O)....	0.27	0.81	0.92	0.17	0.71	0.53	CHEMBL636672
CHEMBL86605 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.009	CC(Cc1ccccc1)Nc2ncnc3	0.21	0.83	0.95	0.17	0.71	0.51	CHEMBL636672
CHEMBL273671 (sdf)	A1	CHEMBL4975	Bos taurus	0.04	CCCNC(=O)N(CCC)c2	0.25	0.83	0.95	0.16	0.71	0.52	CHEMBL640607
CHEMBL273671 (sdf)	A1	CHEMBL4975	Bos taurus	0.05	CCCNC(=O)N(CCC)c2	0.28	0.83	0.95	0.16	0.71	0.53	CHEMBL640605
CHEMBL11002 (sdf)	A1	CHEMBL4975	Bos taurus	0.05	CCCNC(=O)N(CCC)c2	0.17	0.83	0.94	0.14	0.70	0.48	CHEMBL640605
CHEMBL545362 (sdf)	A1	CHEMBL4975	Bos taurus	0.07	CLCCCNC(=O)N(CCC)c2....	0.16	0.82	0.94	0.14	0.71	0.49	CHEMBL640605
CHEMBL11036 (sdf)	A1	CHEMBL4975	Bos taurus	0.08	CCCNC(=O)N(CCC)c2[nH]....	0.14	0.83	0.94	0.14	0.70	0.47	CHEMBL640605
CHEMBL95811 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.14	NCCCC(=O)Nc1nc2ccc(Cl)	0.37	0.83	0.97	0.19	0.72	0.58	CHEMBL849726
CHEMBL158507 (sdf)	A1	CHEMBL4975	Bos taurus	0.15	CCCNC(=O)N(CCC)c2[nH]	0.33	0.85	0.94	0.19	0.71	0.55	CHEMBL640607
CHEMBL415971 (sdf)	A1	CHEMBL4975	Bos taurus	0.15	CCCNC(=O)N(CCC)c2[nH]	0.26	0.84	0.95	0.16	0.71	0.53	CHEMBL640605
CHEMBL85183 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.23	Nc1nc(Cl)nc2c1ncn2C3O....	0.49	0.81	0.91	0.22	0.69	0.61	CHEMBL636672
CHEMBL11163 (sdf)	A1	CHEMBL4975	Bos taurus	0.3	CSc1ccc(cc1)c2nc3C(=O)N(C)....	0.27	0.83	0.94	0.19	0.72	0.54	CHEMBL640607
CHEMBL68738 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.35	OC[C@H]1O[C@H]([C@H](O)	0.34	0.81	0.91	0.17	0.71	0.56	CHEMBL640695
CHEMBL268964 (sdf)	A1	CHEMBL4975	Bos taurus	0.35	CNC(=O)N(C)c2[nH]c	0.30	0.84	0.95	0.18	0.71	0.54	CHEMBL640605
CHEMBL268964 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.4	CNC(=O)N(C)c2[nH]c	0.30	0.84	0.95	0.18	0.71	0.54	CHEMBL642858
CHEMBL9069 (sdf)	A1	CHEMBL4975	Bos taurus	0.4	CNC(=O)Nc2[nH]c(nc2C1=O)	0.35	0.85	0.94	0.23	0.70	0.56	CHEMBL640607
CHEMBL11037 (sdf)	A1	CHEMBL4975	Bos taurus	0.4	CNC(=O)N(C)c2[nH]c(nc2C1=O)....	0.26	0.84	0.95	0.15	0.71	0.53	CHEMBL640607
CHEMBL62350 (sdf)	A1	CHEMBL4975	Bos taurus	0.48	CNC(=O)N(C)c2[nH]c(nc2C1=O)....	0.36	0.85	0.94	0.22	0.71	0.57	CHEMBL640607
CHEMBL11348 (sdf)	A1	CHEMBL4975	Bos taurus	0.48	CCNC(=O)N(CC)c2[nH]c	0.34	0.85	0.94	0.19	0.71	0.56	CHEMBL640607
CHEMBL61845 (sdf)	A1	CHEMBL226	Homo sapiens	0.56	OCC[C@@H]1CCCCN1C(=O)\C=	0.27	0.85	0.95	0.23	0.70	0.52	CHEMBL640343
CHEMBL417607 (sdf)	A1	CHEMBL4975	Bos taurus	0.6	CNC(=O)N(C)c2[nH]c(nc2C1=O)	0.41	0.84	0.95	0.20	0.71	0.59	CHEMBL640607
CHEMBL10873 (sdf)	A1	CHEMBL4975	Bos taurus	0.6	CNC(=O)N(CC=C)C(=O)c2[nH]c	0.32	0.85	0.94	0.32	0.96	0.70	CHEMBL640607
CHEMBL174958 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.7	Cc1nn(C)c2c(O)ncnc12	0.70	0.86	0.95	0.24	0.71	0.70	CHEMBL636669
CHEMBL190 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.7	CNC(=O)N(C)c2nc[nH]c2C1=O	0.76	0.82	0.90	0.32	0.96	0.88	CHEMBL636669
CHEMBL158507 (sdf)	A1	CHEMBL4975	Bos taurus	0.7	CCCNC(=O)N(CCC)c2[nH]c	0.33	0.85	0.94	0.19	0.71	0.55	CHEMBL640605
CHEMBL11104 (sdf)	A1	CHEMBL4975	Bos taurus	0.75	CCCNC(=O)N(CCC)c2[nH]c	0.30	0.85	0.95	0.18	0.71	0.54	CHEMBL640605
CHEMBL269218 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.8	CNC(=O)N(C)c2[nH]c	0.30	0.84	0.94	0.22	0.71	0.54	CHEMBL642858
CHEMBL268026 (sdf)	A1	CHEMBL318	Rattus norvegicus	0.8	CNC(=O)N(C)c2[nH]c	0.30	0.85	0.94	0.19	0.72	0.55	CHEMBL642858
CHEMBL11127 (sdf)	A1	CHEMBL4975	Bos taurus	0.8	CCc1ccc(cc1)c2nc3C(=O)	0.27	0.84	0.94	0.19	0.72	0.54	CHEMBL640605
CHEMBL11410 (sdf)	A1	CHEMBL4975	Bos taurus	0.8	Nc1cc(Cl)ccc1c2nc3C(=O)....	0.27	0.83	0.95	0.20	0.72	0.54	CHEMBL640605
CHEMBL269218 (sdf)	A1	CHEMBL4975	Bos taurus	0.8	CNC(=O)N(C)c2[nH]c	0.30	0.84	0.94	0.22	0.71	0.54	CHEMBL640605

Fig. 4. Tabular results of the similarity search screening performed for the user selected query molecule against all adenosine agonists and antagonists that are deposited in ChEMBL [61]. A similarity based ranked list of a restricted active set of compounds deposited in ChEMBLE is combined with the associated binding data available in literature (1). ChEMBLE IDs (2), receptor target identification (3), as well as several similarity indexes (4) are rapidly user accessible.

mutations at the 6.48 and 6.52 positions that apparently can impair receptor activity, the A_{2A} receptor of *guinea pig* is a totally functional GPCR capable of down-streaming signal upon agonist-induced activation [66]. Furthermore, the comparative sequence analysis of rat and human adenosine A_3 receptors clearly highlights some important differences that could explain the peculiar pharmacological behavior of these two receptors, as summarized in Table S1.

As anticipated, homology modeling remains the most accurate method currently available for predicting the structure of a protein sequence using a homologous template of known structure.

Starting with the first X-ray crystal structure of rhodopsin [67], an increasing amount of information about TM helix packing referring to both active and inactive state of this family of receptors has been collected. The major structural changes associated to agonist binding and receptor activation has been described as an inward tilt of the intracellular part of helix V and an outward tilt of the intracellular portion of helix VI combined with rotation inward tilt of helix VII and an axial shift of helix III [9].

The superimposition of all available A_{2A} receptor crystal structures highlights that the differences in geometrical positions of alpha-carbons of the orthosteric biding pocket residues are modest, characterized by *r.m.s.d.* values comparable to the average resolution of all available crystallographic structures. A comparative analysis carried on the human A_3 adenosine receptor models generating by all possible template all crystallized A_{2A} receptor structures has been performed. After backbones superimposition, slight differences among the relative position of the principal axis of each TM helical domain as well as the geometrical positions of alpha-carbons of the orthosteric biding pocket residues have been detected (see Fig. S2 in the Supplementary Information section). In particular, TM domain VI and VII highlighted the most relevant differences between agonist and antagonist bound structure, which are attributable to Tyr6.27-Ser6.47 segment in TM VI and Asn7.45-Lys7.56 segment in TM VII as shown in Fig. S2.

Several hypothesis of agonist binding based on an inactive GPCR structure have been reported in the literature [68]. This is partially acceptable considering the similar recognition binding motif of both agonist and antagonist in the orthosteric binding site. Despite this fact, several rotameric states of different residues, such as for example Glu169 (ECL2) and His278 (7.43), can guarantee the appropriate complementarity of both agonist and antagonist in the orthosteric binding site. These differences can be relevant in the prediction of a favorable binding mode of agonists and antagonists in particular of novel candidates using molecular docking techniques. A deeper analysis of the accessible volume of the binding site in all human A_{2A} adenosine receptor X-ray structures show a volumetric range from *ca.* 1970 $Å^3$ (PDB ID: 2YDO) to *ca.* 1120 $Å^3$ (PDB ID: 3VGA). Therefore a careful selection of the most appropriate template structure is a mandatory step before starting any receptor-based ligand discovery program. This concept has been also highlighted as a major issue to improving the quality of modeled receptors as clearly demonstrated by the analysis of the results obtained during the past GPCRdock 2008 competition [69].

The development of the "*Best Template Searching*" tool in *Adenosiland* started answering to the question: how can we select the best homologous template to get the highest quality receptor for further molecular docking studies? A possible strategy implemented herein is

Table 1
Comparison of the most relevant bioinformatics information regarding all adenosine receptor subtypes.

	Uniprot ID	Seq. length (AAs)	Id/Sim [%][a]	ECL2 length (AAs)	C-ter length (AAs)	Global *r.m.s.d.* [Cα, Å][b]	TM *r.m.s.d.* [Cα, Å][b]
A1							
Dog	P11616	326	55.4/70.5	35	30	2.12	0.83
Rat	P25099	326	55.1/69.9	35	30	1.44	0.76
Bovine	P28190	326	55.5/70.6	35	30	1.81	0.88
Human	P30542	326	55.9/70.6	35	30	1.08	0.76
Rabbit	P34970	328	55.4/69.5	35	32	1.27	0.87
Guinea Pig	P47745	326	55.1/69.9	35	30	1.80	0.80
Chicken	P49892	324	52.8/70.8	35	28	1.50	0.80
Sumatra orango	Q5RF57	326	54.8/69.9	35	30	1.80	0.83
Mouse	Q60612	326	55.1/69.9	35	30	1.39	0.75
A2A							
Human	P29274	412	100/100	35	116	1.34	0.89
Dog	P11617	412	96.1/97.5	35	116	1.65	0.87
Rat	P30543	410	92.5/96.1	33	119	1.31	0.75
Guinea Pig	P46616	409	90.7/93.6	35	116	1.77	0.94
Mouse	Q60613	410	92.1/96.1	33	110	0.95	0.736
Horse	Q6TLI7	412	93.6/97.2	35	116	0.92	0.77
A2B							
Chicken	O13076	340	72.8/85.8	41	42	1.71	0.83
Human	P29275	332	67.4/82.1	39	34	2.02	0.72
Rat	P29276	332	68.1/82.8	39	34	2.11	0.92
Bovine	Q1LZD0	332	66.1/81.2	38	35	1.62	0.76
Rabbit	Q32ZE2	332	68.9/82.4	39	34	1.95	0.84
Mouse	Q60614	332	69.1/83.5	39	34	1.61	0.92
Dog	Q6W3F4	332	67.8/82.4	38	35	1.34	0.81
A3							
Rabbit	O02667	319	46.2/67.0	29	28	1.86	0.81
Rat	P28647	320	47.9/69.0	29	28	1.26	0.91
Human	P33765	318	47.7/69.2	29	28	1.77	1.02
Sheep	P35342	317	50.0/69.5	28	28	1.25	0.99
Bovine	Q0VC81	317	48.5/70.2	28	28	1.12	0.96
Dog	Q28309	314	49.4/71.3	29	24	1.55	0.97
Mouse	Q61618	319	47.5/68.6	29	28	1.18	0.96

[a] % of similarities and % of identities are calculated using the human A_{2A} primary sequence as reference sequence (UNIProt ID: P29274).
[b] Root mean square deviations (*r.m.s.d.*, in Å) are calculated using the crystallographic structure 3EML [10] as a reference structure.

measure the "similarity" between the new agonist/antagonist and all co-crystallized adenosine ligands. We consider it an interesting example of how bioinformatics bridges chemoinformatics.

3.5. Concluding remarks

Adenosiland represents the first tentative of an integrated bioinformatics and chemoinformatics web-resource dedicated to adenosine receptors. We have already anticipated that *Adenosiland* project can be considered an informatics platform easily transferable to any other GPCR family. In fact, even with several crystal structures available for various GPCRs (some in multiple conformations) it is still not easily possible to predict the exact binding behavior of specific ligands, and it remains difficult to produce crystal structures for certain ligands in complex with their receptors. Therefore, the interdisciplinary approach, presented in this study using both, experimental data and computational predictions, provides valuable information for the rational design of desired highly potent and selective ligands, which are required to validate and exploit their therapeutic potential, and to further elucidate the adenosine receptor's (patho)physiological role. Moreover, the recent evolution of high performance computing infrastructures and, in particular, the application of graphics processing units (GPUs) based molecular dynamics simulations represent important technological innovations that may realize the full potential of atomistic molecular modeling and simulation [70]. As incorporated in *Adenosiland*, the possibility to analyze the pre-equilibrated receptor-membrane systems can be considered a good starting point to appropriately infer to their biophysical behavior in a more realistic micro-environment. In fact, the constantly growing structural information available gave us the opportunity to build up a variety of homology models, giving to *Adenosiland* users the possibility to explore their diversity even from an evolutionary point of view. Moreover users have the possibility to download all human ARs models where template selection is performed using a ligand similarity based approach to the closest ligand already co-crystallized with the human A2A receptor. Indeed, the integrated connectivity with UniProt, RCSB

Table 2
Disulfide bond mapping of adenosine receptor subtypes. All modeled receptors have been constructed following disulfide bridges information reported in the present table.

Receptor subtype	TM3-ECL2 (C77-C166)[a]	ECL2-ECL3 (C74-C146)[a]	ECL2-ECL3 (C71-C159)[a]	TM6-TM7 (C259-C262)[a]
A_1	Yes	No	No	Yes
A_{2A}	Yes	Yes	Yes	Yes
A_{2B}	Yes	No	No	No
A_3	Yes	No	No	No

[a] Numbering is referred to the human A_{2A} subtype considered as reference sequence (UNIProt ID: P29274).

Protein Data Bank, featured in *Adenosiland*, delivers up-to-date information about ARs structural biology and allows scientists to perform similarity search screenings against all adenosine agonists and antagonists deposited in ChEMBLdb, which represents a new important additions in the "state of art" of web based bioinformatics and chemoinformatics platforms.

It is also expected that several implementations will appear over the coming months, in particular following the feedbacks of ver. 1.0 users. In particular, we are integrating in *Adenosiland* all site-directed mutagenesis information collected for all adenosine receptor subtypes as well as an exhaustive depository of the three-dimensional models of all known agonists and antagonists complexes with all human adenosine receptor subtypes, as well as extending the similarity search screening to BindingDB [71].

Acknowledgment

The work was supported by a grant of the Italian Ministry for University and Research (MIUR, FIRB RBNE03YA3L project). The molecular modeling work coordinated by S.M. has been carried out with financial support of the University of Padova, Italy, and the Italian Ministry for University and Research (MIUR), Rome, Italy. S.M. is also very grateful to Chemical Computing Group, YASARA Biosciences GmbH and Acellera for the scientific and technical partnership.

Appendix A. Supplementary material

Supplementary material related to this article can be found at http://dx.doi.org/10.1016/j.ejmech.2012.10.022.

References

[1] G. Burnstock, A. Verkhratsky, Evolutionary origins of the purinergic signalling system, Acta Physiol. (Oxf.) 195 (4) (Apr. 2009) 415–447.
[2] G. Burnstock, Physiology and pathophysiology of purinergic neurotransmission, Physiol. Rev. 87 (2) (Apr. 2007) 659–797.
[3] B.B. Fredholm, A.P. IJzerman, K.A. Jacobson, K.N. Klotz, J. Linden, International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors, Pharmacol. Rev. 53 (4) (Dec. 2001) 527–552.
[4] G. Burnstock, Purinoceptors: ontogeny and phylogeny, Drug Develop. Res. 39 (3–4) (Nov. 1996) 204–242.
[5] E. Dolezelova, M. Zurovec, T. Dolezal, P. Simek, P.J. Bryant, The emerging role of adenosine deaminases in insects, Insect Biochem. Mol. Biol. 35 (5) (May 2005) 381–389.
[6] M.P. Abbracchio, G. Burnstock, A. Verkhratsky, H. Zimmermann, Purinergic signalling in the nervous system: an overview, Trends Neurosci. 32 (1) (Jan. 2009) 19–29.
[7] K.A. Jacobson, Introduction to adenosine receptors as therapeutic targets, Handb. Exp. Pharmacol. (193) (2009) 1–24.
[8] G. Lebon, T. Warne, P.C. Edwards, K. Bennett, C.J. Langmead, A.G.W. Leslie, C.G. Tate, Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation, Nature 474 (7352) (Jun. 2011) 521–525.
[9] F. Xu, H. Wu, V. Katritch, G.W. Han, K.A. Jacobson, Z.-G. Gao, V. Cherezov, R.C. Stevens, Structure of an agonist-bound human A2A adenosine receptor, Science 332 (6027) (Apr. 2011) 322–327.
[10] V.-P. Jaakola, M.T. Griffith, M.A. Hanson, V. Cherezov, E.Y.T. Chien, J.R. Lane, A.P. Ijzerman, R.C. Stevens, The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist, Science 322 (5905) (Nov. 2008) 1211–1217.
[11] A.S. Doré, N. Robertson, J.C. Errey, I. Ng, K. Hollenstein, B. Tehan, E. Hurrell, K. Bennett, M. Congreve, F. Magnani, C.G. Tate, M. Weir, F.H. Marshall, Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine, Structure 19 (9) (Sep. 2011) 1283–1293.
[12] T. Hino, T. Arakawa, H. Iwanari, T. Yurugi-Kobayashi, C. Ikeda-Suno, Y. Nakada-Nakura, O. Kusano-Arai, S. Weyand, T. Shimamura, N. Nomura, A.D. Cameron, T. Kobayashi, T. Hamakubo, S. Iwata, T. Murata, G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody, Nature 482 (7384) (Feb. 2012) 237–240.
[13] W. Liu, E. Chun, A.A. Thompson, P. Chubukov, F. Xu, V. Katritch, G.W. Han, C.B. Roth, L.H. Heitman, A.P. IJzerman, V. Cherezov, R.C. Stevens, Structural basis for allosteric regulation of GPCRs by sodium ions, Science 337 (6091) (Jul. 2012) 232–236.
[14] M. Congreve, S.P. Andrews, A.S. Doré, K. Hollenstein, E. Hurrell, C.J. Langmead, J.S. Mason, I.W. Ng, B. Tehan, A. Zhukov, M. Weir, F.H. Marshall, Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design, J. Med. Chem. 55 (5) (Mar. 2012) 1898–1903.
[15] S. Paoletta, S. Federico, G. Spalluto, S. Moro, Receptor-driven identification of novel human A_3 adenosine receptor antagonists as potential therapeutic agents, Meth. Enzymol. 485 (2010) 225–244.
[16] Apache web server version 2.0 (online). Available at: http://www.apache.org/.
[17] PHP (online). Available at: http://php.net/.
[18] Jmol, the open source molecular viewer (online). Available at: http://jmol.sourceforge.net/.
[19] Jquery Javascript library (online). Available at: http://webworkflow.co.uk/plugins/pfSelect/.
[20] A.M. Waterhouse, J.B. Procter, D.M.A. Martin, M. Clamp, G.J. Barton, Jalview Version 2 – a multiple sequence alignment editor and analysis workbench, Bioinformatics 25 (9) (May 2009) 1189–1191.
[21] M. Clamp, J. Cuff, S.M. Searle, G.J. Barton, The Jalview Java alignment editor, Bioinformatics 20 (3) (Feb. 2004) 426–427.
[22] Yasara (online). Available at: http://yasara.org/.
[23] Molecular operating environment (online). Available at: http://www.chemcomp.com/.
[24] Molecular evolutionary genetics analysis (online). Available at: http://www.megasoftware.net/.
[25] L. Skrabanek, F. Campagne, H. Weinstein, Building protein diagrams on the web with the residue-based diagram editor RbDe, Nucleic Acids Res. 31 (13) (Jul. 2003) 3856–3858.
[26] C. Steinbeck, Y. Han, S. Kuhn, O. Horlacher, E. Luttmann, E. Willighagen, The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics, J. Chem. Inf. Comput. Sci. 43 (2) (Apr. 2003) 493–500.
[27] C. Steinbeck, C. Hoppe, S. Kuhn, M. Floris, R. Guha, E.L. Willighagen, Recent developments of the chemistry development kit (CDK) – an open-source java library for chemo- and bioinformatics, Curr. Pharm. Des. 12 (17) (2006) 2111–2120.
[28] CACTVS (online). Available at: http://www.xemistry.com/.
[29] INDIGO (online). Available at: http://ggasoftware.com/opensource/indigo.
[30] J. Taminau, G. Thijs, H. De Winter, Pharao: pharmacophore alignment and optimization, J. Mol. Graph. Model. 27 (2) (Sep. 2008) 161–169.
[31] M. Floris, J. Masciocchi, M. Fanton, S. Moro, Swimming into peptidomimetic chemical space using pepMMsMIMIC, Nucleic Acids Res. 39 (Web Server issue) (Jul. 2011) W261–W269.
[32] The Universal Protein Resource (UniProt) in 2010, Nucleic Acids Res. 38 (Database issue) (Jan. 2010) D142–D148.
[33] E. Jain, A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B.E. Suzek, M.J. Martin, P. McGarvey, E. Gasteiger, Infrastructure for the life sciences: design and implementation of the UniProt website, BMC Bioinform. 10 (2009) 136.
[34] R. Apweiler, A. Bairoch, C.H. Wu, Protein sequence databases, Curr. Opin. Chem. Biol. 8 (1) (Feb. 2004) 76–80.
[35] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (17) (Sep. 1997) 3389–3402.
[36] J. Qiu, R. Elber, SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs, Proteins 62 (4) (Mar. 2006) 881–891.
[37] A.A. Canutescu, R.L. Dunbrack Jr., Cyclic coordinate descent: a robotics algorithm for protein loop closure, Protein Sci. 12 (5) (May 2003) 963–972.
[38] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack Jr., A graph-theory algorithm for rapid protein side-chain prediction, Protein Sci. 12 (9) (Sep. 2003) 2001–2014.
[39] E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus, Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8, Proteins 77 (Suppl. 9) (2009) 114–122.
[40] A.C. Schiedel, S. Hinz, D. Thimm, F. Sherbiny, T. Borrmann, A. Maass, C.E. Müller, The four cysteine residues in the second extracellular loop of the human adenosine A2B receptor: role in ligand binding and receptor function, Biochem. Pharmacol. 82 (4) (Aug. 2011) 389–399.
[41] P. Labute, Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures, Proteins 75 (1) (Apr. 2009) 187–205.
[42] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (12) (Sep. 2000) 1049–1074.
[43] PovRay, The Persistence of Vision Raytracer (online). Available at: www.povray.org.
[44] M.A. Lomize, A.L. Lomize, I.D. Pogozheva, H.I. Mosberg, OPM: orientations of proteins in membranes database, Bioinformatics 22 (5) (Mar. 2006) 623–625.
[45] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (2) (Jul. 1983) 926–935.
[46] Helmut Grubmüller, VolkerGroll, Solvate (online). Available at: http://www.mpibpc.mpg.de/home/grubmueller/downloads/solvate/index.html.
[47] M.J. Harvey, G. Giupponi, G.D. Fabritiis, ACEMD: accelerating biomolecular dynamics in the microsecond time scale, J. Chem. Theory Comput. 5 (6) (2009) 1632–1639.
[48] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (19) (Nov. 1995) 8577–8593.

[49] V. Kräutler, W.F. van Gunsteren, P.H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, J. Comput. Chem. 22 (5) (2001) 501–508.
[50] MacKerell, D. Bashford, Bellott, Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins†, J. Phys. Chem. B 102 (18) (1998) 3586–3616.
[51] S.C. Lovell, I.W. Davis, W.B. Arendall 3rd, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by Calpha geometry: phi, psi and Cbeta deviation, Proteins 50 (3) (Feb. 2003) 437–450.
[52] B. Vroling, M. Sanders, C. Baakman, A. Borrmann, S. Verhoeven, J. Klomp, L. Oliveira, J. de Vlieg, G. Vriend, GPCRDB: information system for G protein-coupled receptors, Nucleic Acids Res. 39 (Database issue) (Jan. 2011) D309–D319.
[53] C.L. Worth, A. Kreuchwig, G. Kleinau, G. Krause, GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models, BMC Bioinform. 12 (2011) 185.
[54] Y. Zhang, M.E. Devries, J. Skolnick, Structure modeling of all identified G protein-coupled receptors in the human genome, PLoS Comput. Biol. 2 (2) (Feb. 2006) e13.
[55] U. Pieper, B.M. Webb, D.T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E.C. Meng, E.F. Pettersen, C.C. Huang, R.S. Datta, P. Sampathkumar, M.S. Madhusudhan, K. Sjölander, T.E. Ferrin, S.K. Burley, A. Sali, ModBase, a database of annotated comparative protein structure models, and associated resources, Nucleic Acids Res. 39 (Database issue) (Jan. 2011) D465–D474.
[56] F. Kiefer, K. Arnold, M. Künzli, L. Bordoli, T. Schwede, The SWISS-MODEL Repository and associated resources, Nucleic Acids Res. 37 (Database issue) (Jan. 2009) D387–D392.
[57] D. Rodríguez, X. Bello, H. Gutiérrez-de-Terán, Molecular modelling of G Protein-coupled receptors through the web, Mol. Inform. 31 (5) (2012) 334–341.
[58] S. Kelm, J. Shi, C.M. Deane, MEDELLER: homology-based coordinate generation for membrane proteins, Bioinformatics 26 (22) (Nov. 2010) 2833–2840.
[59] S. Jo, T. Kim, V.G. Iyer, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM, J. Comput. Chem. 29 (11) (Aug. 2008) 1859–1865.
[60] P.J. Ballester, W.G. Richards, Ultrafast shape recognition to search compound databases for similar molecular shapes, J. Comput. Chem. 28 (10) (Jul. 2007) 1711–1723.
[61] A. Gaulton, L.J. Bellis, A.P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, Nucleic Acids Res. 40 (D1) (Sep. 2011) D1100–D1107.
[62] I. Gsandtner, M. Freissmuth, A tail of two signals: the C terminus of the A(2A)-adenosine receptor recruits alternative signaling pathways, Mol. Pharmacol. 70 (2) (Aug. 2006) 447–449.
[63] D. Massotte, B.L. Kieffer, The second extracellular loop: a damper for G protein-coupled receptors? Nat. Struct. Mol. Biol. 12 (4) (Apr. 2005) 287–288.
[64] D. Rodríguez, Á. Piñeiro, H. Gutiérrez-de-Terán, Molecular dynamics simulations reveal insights into key structural elements of adenosine receptors, Biochemistry 50 (19) (May 2011) 4194–4208.
[65] V.-P. Jaakola, J.R. Lane, J.Y. Lin, V. Katritch, A.P. Ijzerman, R.C. Stevens, Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: identification and characterization of essential amino acid residues, J. Biol. Chem. 285 (17) (Apr. 2010) 13032–13044.
[66] D. Hirano, Y. Aoki, H. Ogasawara, H. Kodama, I. Waga, C. Sakanaka, T. Shimizu, M. Nakamura, Functional coupling of adenosine A2a receptor to inhibition of the mitogen-activated protein kinase cascade in Chinese hamster ovary cells, Biochem. J. 316 (Pt 1) (May 1996) 81–86.
[67] K. Palczewski, T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, M. Miyano, Crystal structure of rhodopsin: a G protein-coupled receptor, Science 289 (5480) (Aug. 2000) 739–745.
[68] V. Katritch, R. Abagyan, GPCR agonist binding revealed by modeling and crystallography, Trends Pharmacol. Sci. 32 (11) (Nov. 2011) 637–643.
[69] M. Michino, E. Abola, GPCR Dock 2008 Participants, C.L. Brooks, J.S. Dixon, J. Moult, R.C. Stevens, Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008, Nat. Rev. Drug Discov. 8 (6) (Jun. 2009) 455–463.
[70] I. Buch, M.J. Harvey, T. Giorgino, D.P. Anderson, G. De Fabritiis, High-throughput all-atom molecular dynamics simulations using distributed computing, J. Chem. Inf. Model. 50 (3) (Mar. 2010) 397–403.
[71] T. Liu, Y. Lin, X. Wen, R.N. Jorissen, M.K. Gilson, BindingDB: a web-accessible database of experimentally determined protein–ligand binding affinities, Nucleic Acids Res. 35 (Database issue) (Jan. 2007) D198–D201.

2

Investiganting GPCRs-ligand recognition process

Contents

2.1 Modeling the protein-ligand recognition process

Ligand-protein recognition concept is the central topic of Structure Based Drug Design. The challenging issue for the future of drug discovery is the capability to fully understand ligand-protein recognition pathway in order to facilitate the development of drug candidates with more favorable pharmacodynamic profiles. As rule of thumb, good geometric fit and complementarity of hydrophobic and polar contacts between ligand and protein binding site are essential for high biological activity. A rigorous discussion of the statistical thermodynamics of binding have already been reported in literature [59].

Focusing the attention on GPCRs, considering the fact that both ligand and receptors are dynamic entities, recent evidences from functional and biophysical studies supports the existence of multiple receptor conformational states can exist during the entire ligand recognition process [17]. On one hand, homology models represent one of the possible conformation a protein can explore during its lifespan. In such a perspective Hermann Fischer's lock-and-key hypothesis is evolved to a more dynamic model and the conformational selection is thought to happen during orthosteric binding site recognition.
However, the intrinsic mobility of proteins has often been ignored in drug design due to high demand of computational power for its sampling.

Therefore, in this thesis two novel and robust approaches that enable to describe ligand binding at molecular level, has been developed in order to take advantage of Graphical Processing Unit based all-atom Molecular Dynamics simulations in describing complex biomolecular systems.

2.2 Methodological advances

Biomolecular recognition is a complex task to describe at a molecular level. From an experimental point of view, one of the most used technique is isothermal titration calorimetry (ITC)that enables accurate determination of both enthalpic and entropic components of binding [60] [61] [62]. Spectroscopic measurements such as Surface Plasmon Resonance (SPR), Nuclear Magnetic Resonance (NMR) and atomic-force microscopy [63] have shown shown to accurately reproduce binding affinity values that are consistent with ITC [64]. From a computational standpoint ,numerous successful attempts has been developed in order to describe the energetics of binding between a ligand and its protein counterpart. Indeed the common denominator for the thermodynamic description of the binding phenomenon is the study of discrete states of binding. In general, the binding of a ligand with its target protein in aqueous environment can be characterized, at least, by two known discrete states, reported in figure 2.1, such as ligand and protein unbound state and the final complex formation. State functions, such as Gibbs free energy of binding (Δ G), can be used describe accurately this thermodynamic phenomena [59].

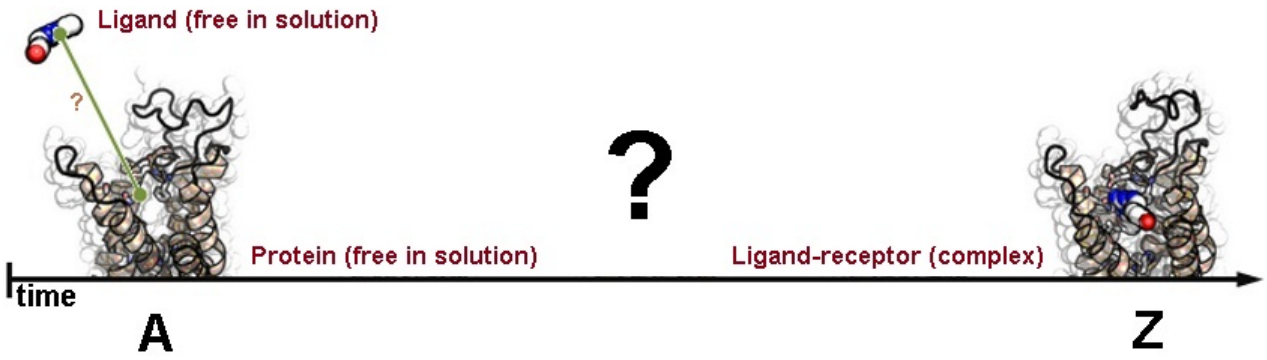


Figure 2.1: *Adenosine Receptor Antagonist ZM241385-human A_{2A} Adenosine Receptor recognition mechanism. Depiction of ligand and protein unbound state (A) and the final complex formation (B). Van der Waals spheres represent ZM241385 atoms and receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). Hydrogen atoms are not displayed.*

All-atom Molecular Dynamics can be used to perform accurate predictions. In details, MM-PBSA (Molecular Mechanics with Poisson-Boltzmann and Surface Area Model) and MM-GBSA (Generalized Born and Surface Area model) are based on the principle that the free energy of binding can be decomposed into individual contribution terms described by molecular mechanic force-fields [65].
Despite the utility of the above cited techniques, the availability of a high resolution ligand-receptor complex obtained by spectroscopical techniques or molecular modeling is required.

In addition, from a pharmacological standpoint, it is very interesting to understand ligand recognition process in much more detailed way. Very few examples in literature are

described of called meta-binding sites characterization using Molecular Dynamics. In facts, this phenomenon is a very rare event to describe at the molecular level and, even with the recent GPU-based [10] or ad-hoc [8] computing resources, it is necessary to carry out classical molecular dynamics experiments in a long microsecond time scale.

In the following chapters we report a methodology that enables to discriminate true binders from an ensemble of decoys thus anticipating the bio-active pose of a ligand.

2.2.1 Molecular Dynamics applicability to Drug Design.

As reported previously, MD simulations that describe spontaneous ligand binding events, without any prior knowledge of the binding site, are computationally intensive. In fact, the major difficult task for molecular dynamics simulations is to assist the design of molecules, with potential therapeutic effect, in time frames that are compatible with experiments.
In the recent years, Shan et al. [66], Dror et al. [8], and Buch et al. [10] performed multiple MD simulations totaling over 150 microseconds, 400 microseconds, and 50 microseconds, respectively. Such investigations regarded the inspection of a single, or a limited number, of binding events even using high performance computing resources that are not normally available for a research group. Despite these efforts, in order to observe a sufficiently high number of binding events to compute the binding affinity of a ligand using unbiased MD simulations, a recent study [18] estimated that seconds to hours of simulated time would be necessary for the purpose. It would take several months to accurately investigate the binding energetics of a single potential pharmacologically active molecule.

The present thesis work highlight a new innovative computational method, named Supervised Molecular Dynamics (SuMD), that allow to follow GPCR-ligand approaching process within a time scale reduced, up to 3 orders of magnitude, compared to classical MD approaches used by other research groups. As reported in figure 2.2 the exploration of the binding pathways of ligands to their receptor counterpart using SuMD accelerate the natural event of binding, which has been reported in past literature to occur in the microsecond time scale.

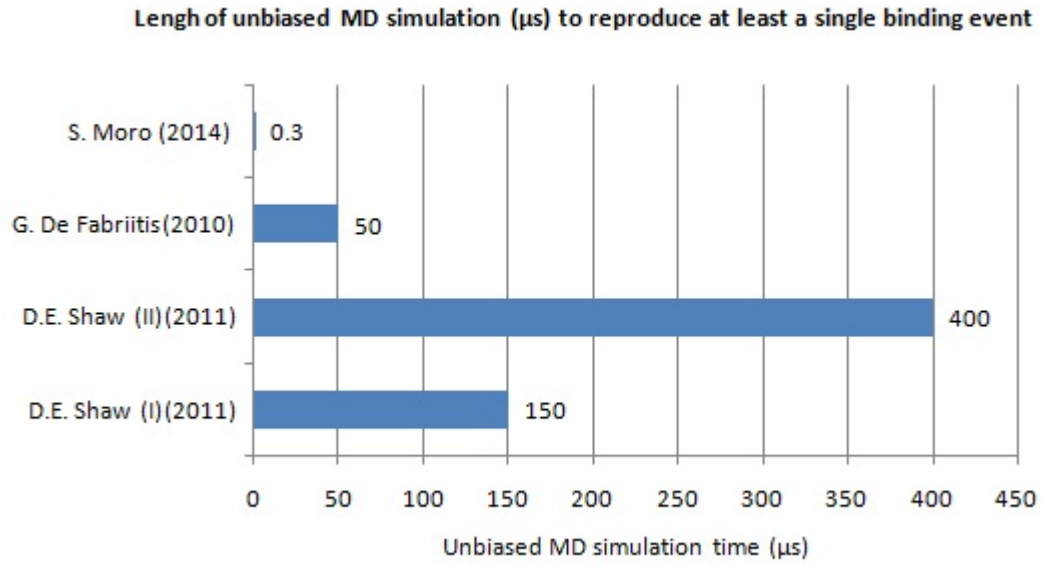


Figure 2.2: *Comparison of the length of unbiased MD simulation time that has to be performed in order to investigate at least a single binding event from different international research groups. D.E. Shaw group study I [66], D.E. Shaw group study II [8] De Fabriitis group study [10] compared to the presented thesis work in S. Moro's group*

This methodological advance, which include an implementation of a tabu-like supervision algorithm on the ligand-receptor distance into the classic Molecular Dynamics (MD) simulation technique, allows to facilitate the characterization of multiple binding events that anticipate the orthosteric binding site and can be applied in a drug design campaign at a high-throughput level in order to design novel binders with preferable pharmacodynamics and kinetics.

In the following chapters the Supervised MD underlying algorithm is reported.

Original publication II

Sabbadin, D., Ciancetta, A., Moro, S. *Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A_{2A} adenosine receptor as a key study* (2014) Journal of Chemical Information and Modeling. In press.

JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

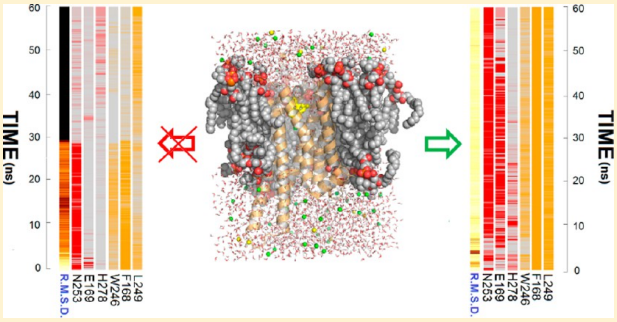
Bridging Molecular Docking to Membrane Molecular Dynamics To Investigate GPCR−Ligand Recognition: The Human A_{2A} Adenosine Receptor as a Key Study

Davide Sabbadin, Antonella Ciancetta, and Stefano Moro*

Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università di Padova, via Marzolo 5, 35131 Padova, Italy

Supporting Information

ABSTRACT: G protein-coupled receptors (GPCRs) represent the largest family of cell-surface receptors and about one-third of the actual targets of clinically used drugs. Following the progress made in the field of GPCRs structural determination, docking-based screening for novel potent and selective ligands is becoming an increasingly adopted strategy in the drug discovery process. However, this methodology is not yet able to anticipate the "bioactive" binding mode and discern it among other conformations. In the present work, we present a novel approach consisting in the integration of molecular docking and membrane MD simulations with the aim to merge the rapid sampling of ligand poses into in the binding site, typical of docking algorithms, with the thermodynamic accuracy of MD simulations in describing, at the molecular level, the stability a GPCR-ligand complex embedded into explicit lipid−water environment. To validate our approach, we have chosen as a key study the human A_{2A} adenosine receptor (hA_{2A} AR) and selected four receptor−antagonist complexes and one receptor−agonist complex that have been recently crystallized. In light of the obtained results, we believe that our novel strategy can be extended to other GPCRs and might represent a valuable tool to anticipate the "bioactive" conformation of high-affinity ligands.



INTRODUCTION

G protein-coupled receptors (GPCRs) are the largest family of cell-surface receptors and represent ∼3% of the genes in the human genome.[1] They regulate several crucial functions of most cells in the body, and receptor dysfunction can lead to a variety of disease conditions.[2] These receptors respond to a wide variety of structurally diverse ligands, ranging from small molecules (such as biogenic amines, nucleotides, and ions) to lipids, peptides, proteins, and even light.[3] Ligands (agonists, inverse agonists, and antagonists) acting on GPCRs play an important role in the treatment of numerous diseases, including cardiovascular and mental disorders, cancer, and viral infections.[2] It is estimated that these receptors represent about one-third of the actual identified targets of clinically used drugs.[4,5] The determination of the rhodopsin crystal structure and, more recently, adrenergic, dopaminergic, histaminergic, opioid and A_{2A} adenosine receptors provides both academia and pharma companies with exceptionally valuable information for a better understanding of the molecular determinants of receptor function and a more-reliable rationale for drug design.[6]

The progress made in the field of GPCRs structural determination has increased the adoption of docking-based screening for novel potent and selective ligands with a potentially significant savings of time and money. However, despite many advances carried out in the molecular docking field during the past decade, this methodology is still far from being realistic and accurate.[7] More commonly, the goodness of chemical complementarity between the ligand and its receptor is evaluated by an energy function (*scoring function*) composed of different energetic terms that attempt to account for the *forces driving* ligand *binding* to the receptor.[8] As recently demonstrated, docking programs are usually successful in generating multiple poses that include binding modes similar to the crystallographically determined bound structure, whereas scoring functions are much less successful at correctly identifying the "bioactive" binding mode.[7] This narrows the applicability of the methodology to those cases where the crystallographic structure is available for comparison and generally implies the need for the calibration of the docking protocol through benchmark studies.

However, to date, only ∼1% of GPCRs structures has been experimentally determined with the consequence that the research focused on the majority of the targets of interest is based on structures obtained by homology modeling. Therefore, novel approaches are needed to increase docking robustness and applicability, not only to anticipate the "bioactive" pose of a ligand within the receptor crystallographic structure but also to discriminate true binders from an ensemble of decoys.

Received: September 23, 2013

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

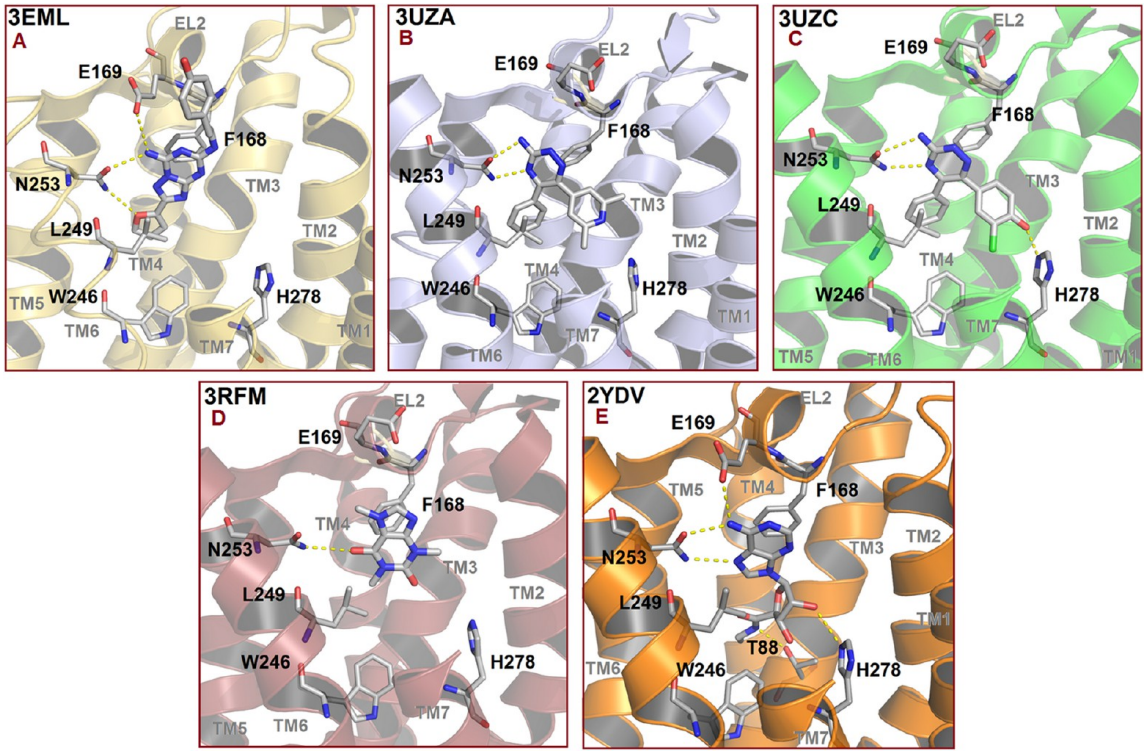


Figure 1. Overview of the binding modes at the hA_{2A} AR of the selected co-crystallized ligands: (A) ZM 241385 (PDB ID: 3EML), (B) T4G (PDB ID: 3UZA), (C) T4E (PDB ID: 3UZC), (D) caffeine (PDB ID: 3RFM), and (E) NECA (PDB ID: 2YDV). The complexes are viewed from the membrane side facing TM6 and TM7, with the view of TM7 partially omitted. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed, whereas hydrogen bond interactions are highlighted as yellow dashed lines.

Very recently, sophisticated molecular dynamics (MD) approaches adapted to massively parallel computer architectures have allowed the execution of microsecond-scale standard MD simulations of fully atomistic representations of GPCRs embedded into explicit lipid–water environments.[9] Unfortunately, although MD simulations represent the highly accurate modeling methods dealing with macromolecular systems, they still remain computationally expensive and require costly high-performance computing (HPC) resources.[10] The recent evolution of commodity graphics processing units (GPUs) represents an important technological innovation that may realize the full potential of atomistic molecular modeling and simulation.[11] In fact, the ability to rapidly compute realistic estimates of binding energies would be of great use in drug discovery process in particular as a robust alternative to the conventional scoring functions in molecular docking field. However, the exact thermodynamic methods using MD require long-running simulations, thus making the total computation time uncompetitive with direct experimental measurements. To be a practical complement to the experimental techniques, a computational method must have a time-to-answer on the order of a few days. Interestingly, GPUs can currently deliver more floating-point operations per second (FLOPS) (by more than an order of magnitude) than standard processors, thereby also drastically reducing the computational time cost of long-term fully atomistic MD simulations.[11]

In the present work, we present a novel approach consisting in the integration of molecular docking and membrane MD simulations with the aim to merge the main advantage of docking, that is the rapid sampling of ligand poses into the binding site, with the thermodynamic accuracy of MD simulations in particular regarding the description, at the molecular level, of the stability a GPCR–ligand complex embedded into explicit lipid–water environment. In other words, we would like to verify if the "energy inspection" of membrane MD trajectory obtained starting from different GPCR–ligand docking poses may improve our ability to identify the "bioactive" pose of a ligand within the receptor crystallographic structure. This would allow us to overcome some of the most crucial "energy-related" approximations of the conventional scoring functions, such as the absence of explicit water molecules and the exploration of the GPCR-ligand complex flexibility. In particular, all membrane MD simulations have been carried out using an ACEMD program engineered to run on GPUs.[12]

To validate our combined approach, we have selected, as a key study, the human A_{2A} adenosine receptor (hA_{2A} AR), which has been recently crystallized with several ligands, both agonists and antagonists, characterized by different receptor binding affinities. In particular, we have focused our attention on hA_{2A} AR antagonists since they are gaining interest because of their potential use for the treatment of a variety of neurological disorders, such as Parkinson's disease, Huntington's disease, and migraines.[13] Recently, phase III studies on Preladenant, which is an A_{2A} AR antagonist, as a potential drug for the treatment of Parkinson's disease were abandoned thus proving the need for concrete approaches that are able to improve the quality of GPCRs–ligand models for docking and screening applications and enable a detailed structural investigation of GPCRs–ligand interaction, by taking into account the role of water molecules in ligand binding as well as the influence of the membrane on protein flexibility.

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

For the present study, we have selected five crystal structures of the hA_{2A} AR in complex with four strong binders, such as 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino)ethyl)phenol, ZM 241385 (pK_D = 9.18 ± 0.00,[14] Protein DataBank (PDB) ID: 3EML[15]); 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pK_D = 8.9 ± n.d.,[14] PDB ID: 3UZA[16]); 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pK_D = 9.6 ± n.d.,[14] PDB ID: 3UZC[16]); NECA, *N*-ethyl-5′-carboxamido adenosine (pK_D = 7.00 ± 0.1,[14] PDB ID: 2YDV[17]) and a weaker binder such as caffeine (pK_D = 5.31 ± 0.44,[14] PDB ID: 3RFM[14]). The structures of the considered antagonists inside the orthosteric binding pocket of the hA_{2A} AR are collected in Figure 1. The obtained results prove that the strategy is able to reproduce the "bioactive" conformation of high affinity ligands and to discern it among other "less stable" conformations, as described in details in the following.

METHODS

Computational Facilities. All computations were performed on a hybrid CPU/GPU cluster. In particular, molecular docking simulations have been carried out using 8 Intel Xeon E5620 CPU cluster, whereas membrane molecular dynamics simulation have been performed with a 4 NVIDIA GTX 580 and 2 NVIDIA GTX 680 GPU cluster engineered by Acellera.[18]

In the following, the numbering of the amino acids follows the arbitrary scheme by Ballesteros and Weinstein: each amino acid identifier starts with the helix number, followed by the position relative to a reference residue among the most conserved amino acids in that helix, to which the number 50 is arbitrarily assigned.[19]

Homology Models. The selected five crystal structures (PDB IDs: 3EML, 3UZA, 3UZC, 3RFM, and 2YDV) and the FASTA sequence of the hA_{2A} AR (Uniprot ID: P29274) were retrieved from the RCSB PDB database[20] (http://www.rcsb.org) and the UniProtKB/Swiss-Prot,[21−23] respectively. The eventual lysozyme portion fused to the receptor, as well as co-crystallized ligands and water molecules, have been removed before starting the homology modeling procedure. Ionization states and hydrogen positions have been assigned with the "Protonate-3D" tools.[24] Then, to minimize contacts among hydrogens, the structures were subjected to energy minimization with Amber99 force field[25] until the root-mean-square (RMS) of the conjugate gradient was <0.05 kcal mol^{-1} $Å^{-1}$, by keeping the heavy atoms fixed at their crystallographic positions. The FASTA sequence was aligned, using Blosum 62 matrix, with the template sequence. Backbone and conserved residues coordinates were copied from the template structure, whereas newly modeled regions and nonconserved residues side chains were modeled and energetically optimized using an Amber99 force field until a RMS of the conjugate gradient of <0.05 kcal mol^{-1} $Å^{-1}$ was reached. Missing loop domains were constructed by the loop search method implemented in the Molecular Operating Environment (MOE, version 2010.10) program,[26] on the basis of the structure of compatible fragments found in the PDB.[20] N-terminal and C-terminal were deleted if their lengths exceeded those found in the crystallographic template. The "Protonate-3D" tool was used to appropriately assign ionization states and hydrogen positions[24] to the build models. Protein stereochemistry evaluation was then performed by employing several tools (Ramachandran and χ plots measure j/ψ and χ_1/χ_2 angles, clash contacts reports) implemented in the MOE suite.[26]

Docking. Co-crystallized agonist and antagonists structures were extracted from the original protein−ligand complex coordinates files and checked for errors. Hydrogen atoms were added and the protonation state (pH 7.4) was checked. Partial charges for ligands were imported from the MOPAC program output files using the PM3/ESP semiempirical Hamiltonian,[27,28] whereas partial charges for protein amino acids were calculated on the basis of the Amber99 force field. Ligands were docked into the orthosteric binding site of the hA_{2A} AR models with the GOLD 5.1 suite using the genetic algorithm protocol[29] (10 independent docking runs for each compound) and the CHEMPLP scoring function. The outcoming poses have been then rescored on the basis of the GoldScore scoring function.[29,30] The latter, in a previous study,[31] resulted the best among the tested scoring functions in reproducing and ranking the crystallographic binding mode of ZM 241385 at the hA_{2A} AR.[32] The purpose of the docking procedure was to use the search algorithm to identify as many different binding modes as possible: we therefore forced the program to retain 10 poses that differed in terms of the root-mean-square deviation (RMSD) for at least 1.75 Å, by setting the non default "diverse solutions" keyword, as implemented in the GOLD suite. The resulting conformations have been sorted according to the cluster number. The values of the Fitness Score, as evaluated by the GoldScore scoring functions, and the RMSD values with respect to the corresponding crystal structures are reported in Table S1 in the Supporting Information.

Interaction Energy Fingerprints (IEFs). To analyze the ligand−receptor recognition mechanism in a more quantitative manner, we calculated the individual electrostatic and hydrophobic contributions to the interaction energy (hereby denoted as IE_{ele} and IE_{hyd}, respectively) of each receptor residue involved in the binding with the ligand. In particular, the electrostatic contribution has been computed on the basis of the nonbonded electrostatic interaction energy term of the force field,[33] whereas the hydrophobic contributions has been calculated by using the directional hydrophobic interaction term based on contact surfaces as implemented in the MOE scoring function.[26] As a consequence, an energy (expressed in units of kcal mol^{-1}) is associated to the electrostatic contribution, whereas a score (the higher the better) is related to the hydrophobic contribution.

The analysis of these contributions have been reported as "interaction energy fingerprints" (hereby indicated as IEFs), i.e., interaction energy patterns (graphically displayed either as histograms or as heatlike maps) reporting the key residues involved in the binding with the considered ligands along with a quantitative estimate of the occurring interactions.

Molecular Dynamics. Each ligand−receptor complex was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer (75 Å × 75 Å wide) and placed into the membrane according to the suggested orientation reported in the "Orientations of Proteins in Membranes (OPM)" database[34] for the hA_{2A} AR in complex with the antagonist T4G (PDB ID: 3UZA[16]). The membrane has been generated by using a grid-based method[35] with the VMD Membrane Plugin tool[36] and overlapping lipids (within 0.6 Å) were removed upon insertion of the protein. The total number of lipids composing the lipid bilayer of each considered membrane-embedded ligand−protein system are reported in Table S2 in the Supporting Information (upper panel), whereas a detailed

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

representation is depicted in Figure S2 in the Supporting Information (panel I).

The prepared systems were solvated with TIP3P[37] water using the Solvate 1.0 program[38] and neutralized by Na^+/Cl^- counterions to a final concentration of 0.154 M. The total number of atoms per system was ~35 000. Membrane MD simulations were carried out on a GPU cluster with the ACEMD program,[12] using the CHARMM27 Force Field[33] and periodic boundary conditions. Initial parameters for the ligands were derived from the CHARMM General Force Field for organic molecules,[39] using the "paramchem" service,[40,41] and were subsequently optimized at the MP2/6-31G* level of theory[42] (which is consistent with the CHARMM27 Force Field parametrization) using Gaussian 09[43] and the implemented parametrization tools in the VMD engine.[36]

The system was equilibrated using a stepwise procedure. In the first stage, to reduce steric clashes due to the manual setting up of the membrane-receptor system, a 500-step conjugate-gradient minimization was performed. Then, to allow lipids to reach equilibrium and water molecules to diffuse into the protein cavity, the system was equilibrated by keeping the positions of protein and ligand atoms restrained for the first 8 ns, using a force constant of 1 kcal mol^{-1} $Å^{-2}$ and then by keeping only the α carbon atoms frozen up to 9 ns while gradually reducing the force constant to 0.1 kcal mol^{-1} $Å^{-2}$. During the equilibration procedure, the temperature was maintained at 298 K, using a Langevin thermostat with a low damping constant of 1 ps^{-1}, and the pressure was maintained at 1 atm using a Berendensen barostat. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE[44] algorithm with an integration time step of 2 fs.

In order to assess the biophysical validity of the built systems, the average area per lipid headgroup (APL) and bilayer thickness measurements for each built system was measured using Grid-MAT-MD.[45] The corresponding averaged area per lipid headgroup of the extracellular leaflet (eAPL) and of the intracellular leaflet (iAPL) in the first nanosecond (eAPL1 ns and iAPL1 ns) and in the last nanosecond (eAPL9 ns and iAPL9 ns) of the equilibration for all the considered complexes is reported in Table S2 in the Supporting Information (lower panel). The calculated values are in agreement with the experimental values measured for 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayers.[46] Bilayer system representation and the performed thickness analysis, for each built system at the end of the equilibration phase, are reported in Figure S2 in the Supporting Information (panels I and II, respectively). Harmonical constraints were then removed during an additional 60 ns (NVT ensemble). Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)[47] with grid size rounded to the approximate integer value of cell wall dimensions. A nonbonded cutoff distance of 9 Å with a switching distance of 7.5 Å was used.

The total of 69 ns of membrane molecular dynamics took ~45 h of NVIDIA GTX580 GPU time per trajectory. All molecular dynamics experiments were carried out in triplicate for a total of ~10 μs of MD trajectories that resulted in ~1100 h per single GPU used.

Dynamic Scoring Function. The dynamic scoring function (DSF) is defined as the cumulative sum of the electrostatic (IE_{ele}) and hydrophobic (IE_{hyd}) contributions to ligand binding during the MD trajectories computed at frames extracted every 100 ps. To calculate such contributions, dynamic selections of residues within a range of 4.5 Å from the ligand have been selected for the calculation of the electrostatic DSF (DSF_{ele}, eq 1) and the hydrophobic DSF (DSF_{hyd}, eq 2):

$$DSF_{ele} = \sum_{t=0}^{n} IE_{ele} \quad (1)$$

$$DSF_{hyd} = \sum_{t=0}^{n} IE_{hyd} \quad (2)$$

Moreover, to take into account the degree of fitness of the predicted binding conformations and to highlight differences between stable and unstable poses, we also calculated the value of the weighted DSF (wDSF) by dividing the values in eqs 1 and 2 by the ligand fluctuation (RMSD), with respect to the starting position generated by the docking protocol. The corresponding weighted electrostatic and hydrophobic DSFs (denoted as $wDSF_{ele}$ and $wDSF_{hyd}$, respectively) are reported in eqs 3 and 4:

$$wDSF_{ele} = \frac{\sum_{t=0}^{n} IE_{ele}}{RMSD} \quad (3)$$

$$wDSF_{hyd} = \frac{\sum_{t=0}^{n} IE_{hyd}}{RMSD} \quad (4)$$

The obtained DSF and wDSF values then were plotted against the simulation time and generic linear functions ($f(x) = m{\cdot}x$) were fitted to the collected data. The slope coefficients of the fitted lines provide an estimate of the enduring strength of the interaction with the nearby residues, thus highlighting differences between stable and unstable binding modes: Higher slope coefficients (absolute value) are associated with ligand conformations that are strongly and steadily bound to the residues, whereas lower slope values correspond to ligand conformations that possess a low degree of fitness with the binding pocket and are expected to differ from the "bioactive" conformation. Slope coefficients are reported in Table S3 in the Supporting Information.

Multimedia Materials. Trajectory analysis and the generation of figures and videos were performed using several functionalities implemented in VMD,[36] the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC, and a Gnuplot graphic utility (http://www.gnuplot.info/).

RESULTS AND DISCUSSION

General Features of the Orthosteric Binding Site of the hA_{2A} AR. The binding site of the hA_{2A} AR has been exhaustively described elsewhere.[6] Therefore, here, we report the most relevant receptor−ligand binding features that we have taken into account to inspect and analyze the results of molecular docking and MD simulations. As depicted in Figure 1, the common interaction pattern for all ligands involves an aromatic $\pi-\pi$ stacking with the conserved Phe168, located in the second extracellular loop (EL2), and additional hydrophobic contacts with, among others, the Leu249 (6.51) side chain. Strong polar interactions are established with the side chain of the conserved Asn253 (6.55),[48] where the role of the hydrogen bond donor in the high-affinity ligands is played by an exocyclic amine group. In the structure co-crystallized with ZM 241385 (Figure 1A), the side chain of Glu169 (EL2) is involved in an additional hydrogen bond, whereas in the other

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

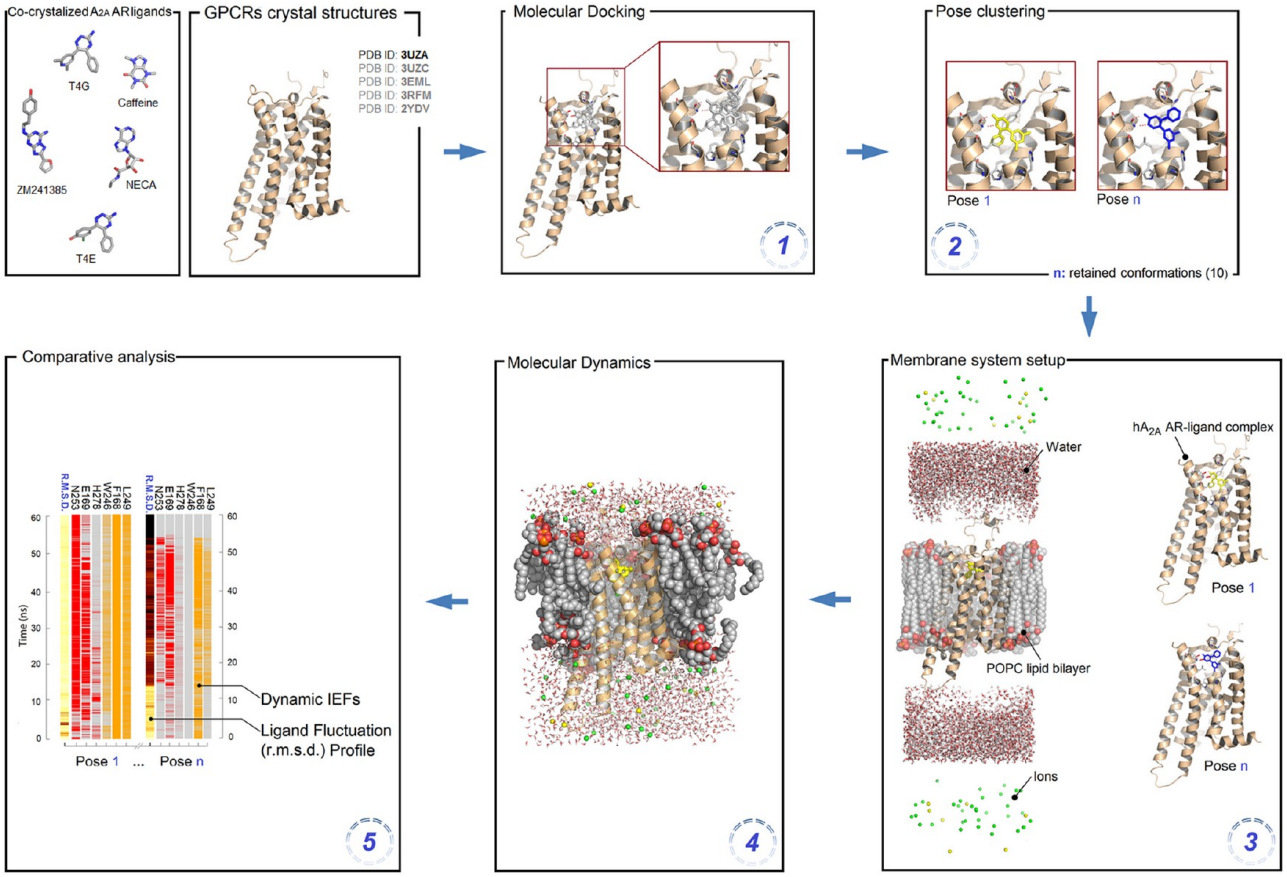


Figure 2. Workflow of the combined molecular docking and membrane molecular dynamics protocol.

structures, the residue is found in a different rotameric state, preventing such interaction. Moreover, in the agonist-bound crystal structure, the Thr88 (3.36) side chain forms a hydrogen bond interaction with the nitrogen atom of the acetamide moiety in NECA. This pattern is consistent with the previously reported mutation data, which have been recently reviewed by Crystalli and collaborators,[49] showing a loss of affinity for the Asn253 (6.55) mutant, as well as with recent mutagenesis data,[48] revealing the critical role of Phe168 (EL2) and Leu249 (6.51) for both agonists and antagonists binding and of Thr88 (3.36) for agonist binding.

Workflow of the Combined Molecular Docking and Membrane MD Protocol. As anticipated in the Introduction, one of the most difficult tasks in structure-based drug discovery is the accurate prediction of receptor−ligand binding interactions. For this purpose, molecular docking and scoring functions are the most used approaches. However, often, the top-ranked docking poses do not represent the "bioactive" (crystallographic) binding mode, and very frequently, there is no correlation between docking scores and binding affinity data.[50,51] Therefore, the "post-processing" of docking poses has recently emerged as a strategy to raise the success of docking studies and several approaches have been proposed.[52] In this work, we present an alternative method consisting in the integration of molecular docking and membrane MD simulations with the aim to merge the main advantage of docking, that is the rapid sampling of ligand poses into in the binding site, with the thermodynamic accuracy of MD simulations in particular regarding the description, at the molecular level, of the stability a GPCR-ligand complex embedded into explicit lipid−water environment. The workflow of the combined protocol is shown in Figure 2: Starting from a conventional receptor-driven docking protocol, the top 10 ranked poses have been clustered (clustering distance = 1.75 Å) and each pose-receptor complex was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer and subjected to 60 ns of MD simulations (in triplicate). From the resulting MD trajectories, we analyzed in details the following three aspects: (i) the evolution of the IEFs (the hereby denoted as "dynamic IEFs") that highlights if the interaction of the ligands with the surrounding residues is conserved throughout the considered time lapse, (ii) the ligand fluctuation profile (expressed in terms of RMSD) that reflects the "positional stability" of the starting conformation, and (iii) the cumulative sum of ligand receptor interactions that provides a dynamic estimate of both the positional stability and the strength of the interaction network. Indeed, as in principle, the "bioactive" (crystallographic) binding mode is the one in which an high-affinity ligand is strongly anchored to its orthosteric binding site, it is expected that the docking pose that better reproduces it shows both a stable position and a persistent interaction network during the simulations. To validate our protocol, we have selected five crystal structures of the hA_{2A} AR in complex with four strong binders such as ZM 241385 (PDB ID: 3EML), T4G (PDB ID: 3UZA), T4E (PDB ID: 3UZC), NECA (PDB ID: 2YDV), and the weaker binder caffeine (PDB ID: 3RFM).

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

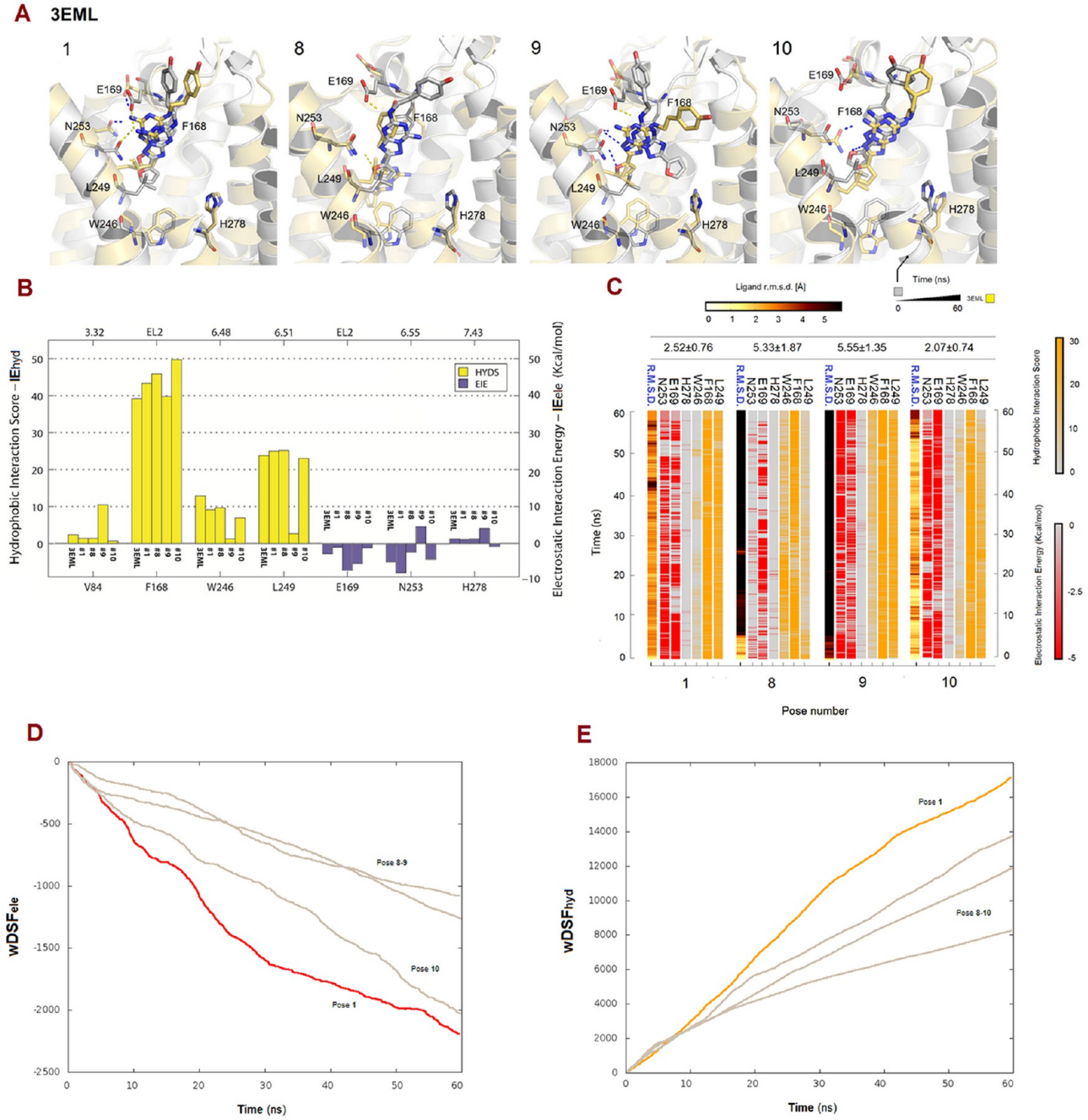


Figure 3. (A) Docking poses of ZM 241385 at the hA_{2A} AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic wDSFs and (E) hydrophobic wDSFs. IE_{ele} values are given in units of kcal $Å^{-1}$ mol^{-1}, IE_{hyd} values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.

In Silico Inspection of ZM 241385 Binding Mode. We first tested our combined procedure with the potent hA_{2A} AR antagonist ZM 241385: In the majority of the retained docking poses (poses 1−7), the aromatic core is superimposed (poses 1−7: RMSD < 1 Å; see Table S1 in the Supporting Information and Figure 3A) to the corresponding co-crystallized ligand and the observed interaction patterns reflect the same key interactions highlighted by the crystallographic structure[15] (Figure 3B).

Membrane MD simulation of pose 1 (see Figure 3A, as well as Table S1 in the Supporting Information) is characterized by a high positional stability (RMSD = 2.52 ± 0.76 Å) and the analysis of dynamical IEFs (see Figure 3C) reveals strong and persistent interactions with Asn253 (6.55), Glu169 (EL2), Phe168 (EL2), and Leu249 (6.51). Similar results have been obtained from the MD simulations of all of the other seven ligand-protein complexes (poses 2−7, data not shown). Although the IEFs pattern of pose 8 (Figure 3B) apparently mirrors that observed for the co-crystallized structure, the ligand has a different orientation into the binding pocket. Indeed, the exocyclic amino group points toward TM2 and does not interact with any residue within a range of 4.5 Å, and

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

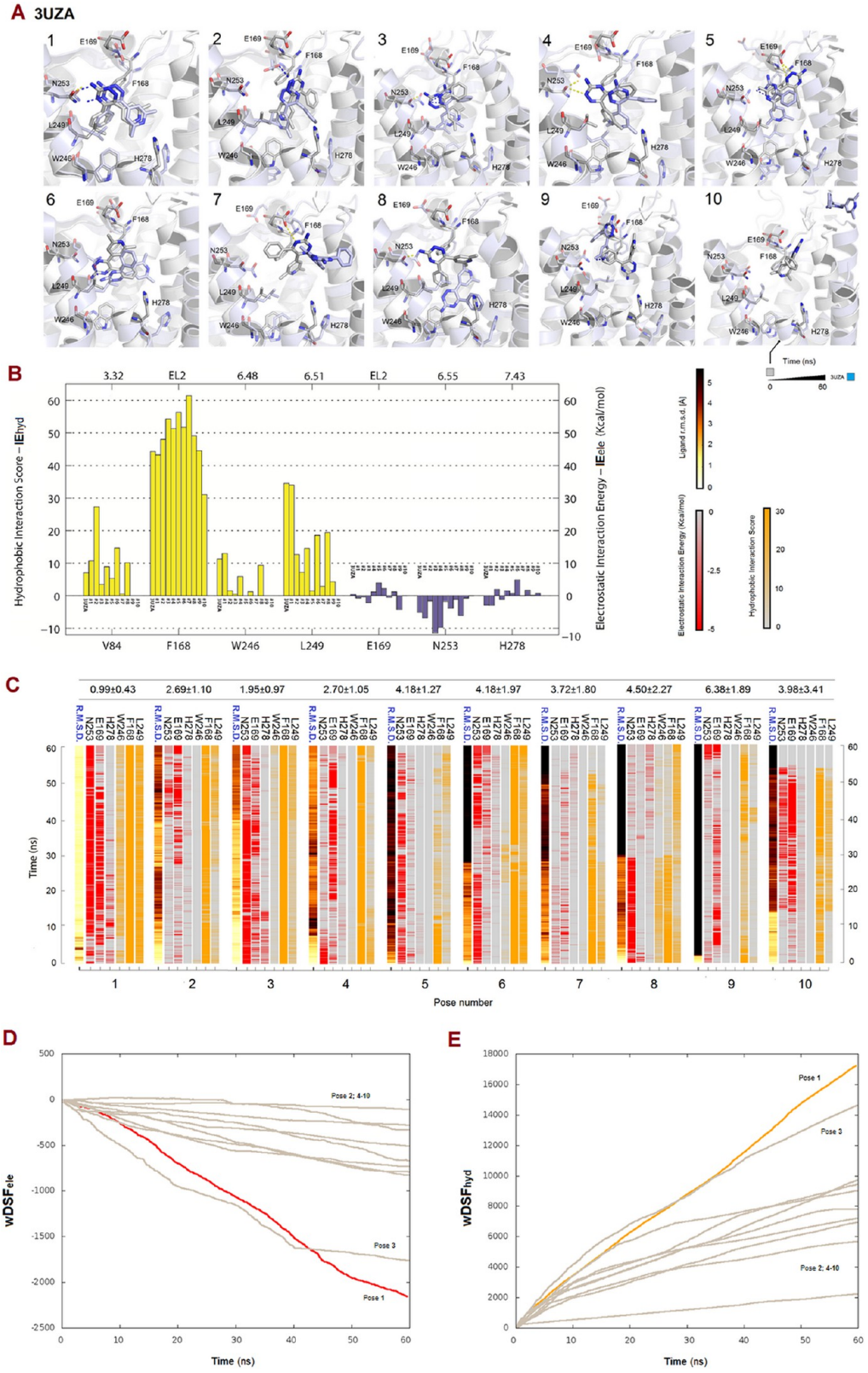


Figure 4. (A) Docking poses of T4G at the hA_{2A} AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic wDSFs and (E) hydrophobic wDSFs. IE_{ele} values are given in units of kcal $Å^{-1}$ mol^{-1}, IE_{hyd} values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

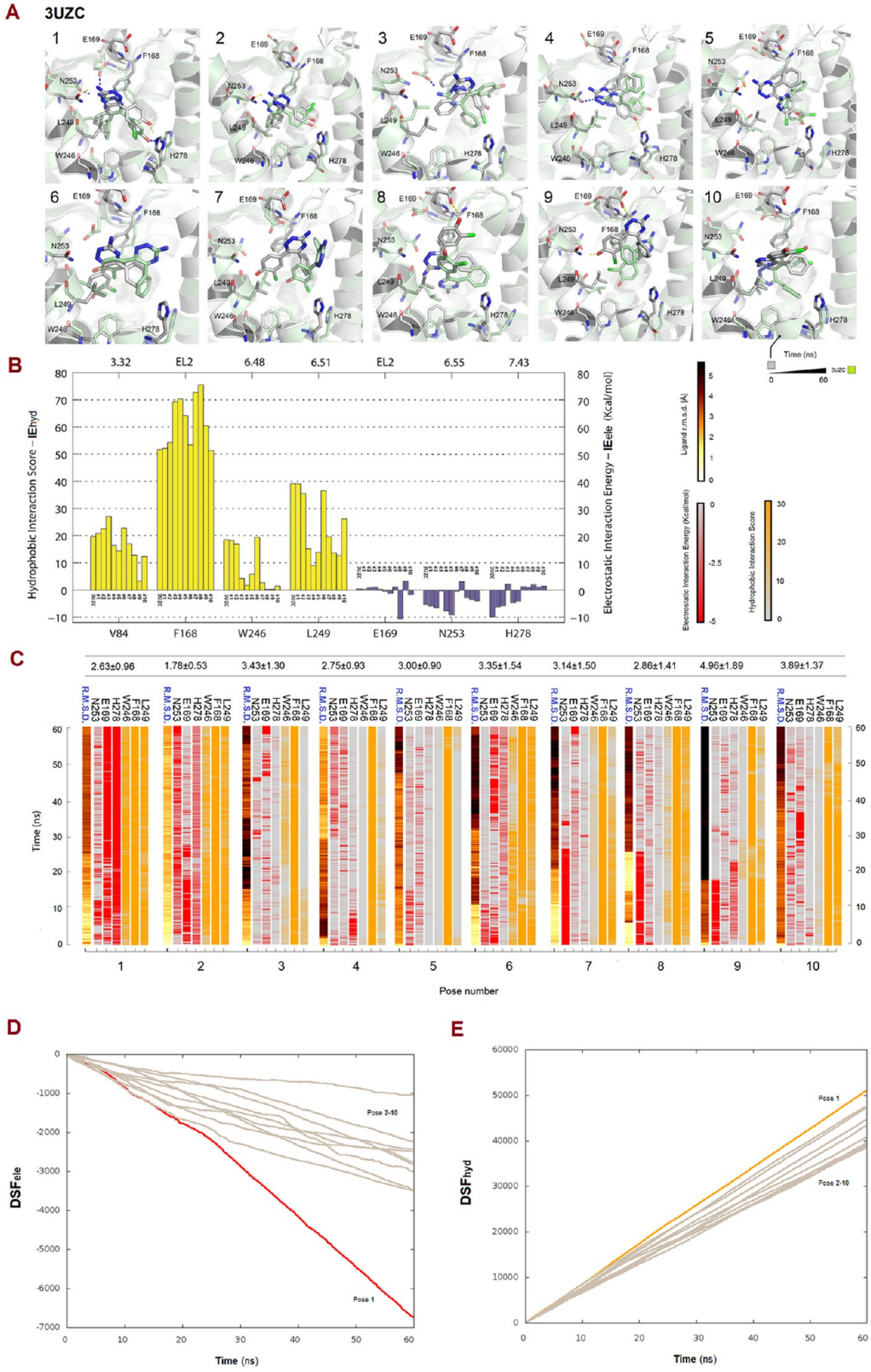


Figure 5. (A) Docking poses of T4E at the hA_{2A} AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic DSFs and (E) hydrophobic DSFs. IE_{ele} values are given in kcal/mol, IE_{hyd} values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the polar interactions with Asn253 and Glu169 are established by the nitrogen atom in the triazole ring and the aminoethyl moiety, respectively (see Figures 3A and 3B). During the MD simulation (Figure 3C), these initially predicted strong hydrogen bonds are lost and the high average RMSD value (5.33 ± 1.87 Å) further confirms the low degree of fitness of the starting pose. Conversely, although poses 9 and 10 also exhibit a different initial orientation of the scaffold, with respect to the co-crystallized ligand (RMSD: 3.58 and 2.47 Å, respectively), during the MD simulation both poses are able to establish strong polar interactions with Asn253 (6.55) and Glu169 (EL2) (see Figures 3A and 3C).

Moreover, as reported in Table S3 in the Supporting Information and graphically displayed in Figures 3D and 3E, the slope coefficients fitted on the DSFs and wDSFs highlight that the conformations that are nearly superimposable to the crystallographic information (pose 1) or whose evolution, during the MD simulation, converge to the crystallographic conformation[15] (pose 10) have the more favorable slope absolute values.

These results suggest new insights into the late recognition process of ZM 241385 at the hA_{2A} AR, an aspect which has been extensively and uniquely described for the β_2-adrenergic receptor.[9] The existence of possible meta-binding conformations, identified by poses 9 and 10 (Figure 3A), enriches the description of the events that might occur once the ligand has entered the binding cavity. More studies (e.g., by employing nonequilibrium MD methods such as steered MD) are needed to better assess the statistical probability of these events and to clarify the role of small energy barriers among different ZM 241385 binding conformations, recently detected by X-ray crystallography.[14]

In Silico Inspection of T4G Binding Mode. The conformations of T4G (Figure 4A) inside the orthosteric binding site of the hA_{2A} AR have RMSD values, with respect to the crystallographic structure, that span from 0.69 Å (pose 1, see Table S1 in the Supporting Information) to 7.72 Å (pose 10, see Table S1 in the Supporting Information). The variability of the conformations is also reflected by the corresponding IEFs, as depicted in Figure 4B.

Pose 1 (Figure 4A) shows an interaction pattern involving residues that play a critical role in antagonist recognition (Figure 4B): Both the endocyclic and exocyclic nitrogen atoms of the aromatic scaffold establish hydrogen bonds with Asn253 (6.55). The 5-phenyl ring is directed toward the conserved His250 (6.52) and Trp246 (6.48) residues and interacts with the hydrophobic side chain of Val84 (3.32). A $\pi-\pi$ stacking interaction occurs between the conserved Phe168 (EL2) side chain and the 1,2,4-triazine aromatic core, which additionally interacts with Leu249 (6.51). During the MD simulation (Figure 4C), the polar contacts with Asn253 (6.55) are maintained strong and persistent. Interestingly, Glu169 (EL2), which is not initially involved in any interaction with the ligand (Figure 4B), is recruited during the simulation and establishes an hydrogen bond with the exocyclic nitrogen atom (Figure 4A). The dynamic IEFs (Figure 4C) also highlight favorable and persistent hydrophobic contacts with Leu249 (6.51), Phe168 (EL2), and Trp246 (6.48). The stability of the initial binding mode is also confirmed by the low ligand fluctuation (average RMSD = 0.99 ± 0.43 Å). Again, also in the case of T4G, among the other poses, there are some whose interaction patterns share common features with the co-crystallized ligand (poses 3, 4, 5, 7, and 8) and others that do not establish polar interactions with key residues involved in antagonist recognition (poses 2, 6, 9, and 10). In all these cases, the MD simulations (Figure 4C) have revealed unstable interaction patterns and a low positional stability with a consistent increase of ligand fluctuations into the binding site.

This scenario is further confirmed by the analysis of the DSFs (Figures 4D and 4E) and their corresponding slope coefficients (Table S3 in the Supporting Information): The higher slope correspond to pose 1, whereas lower slopes are associated to the other possible binding conformations. The difference can be graphically detected in Figures 4D and 4E.

In Silico Inspection of T4E Binding Mode. As for T4G, also the docking poses for T4E (Figure 5A) show a variable range of superimposition with respect to the crystallographic structure ranging from 0.33 Å to 6.58 Å (Table S1 in the Supporting Information) with associated docking scores comprising bad placements (pose 10) as well as poses to which a high score is assigned (poses 1 and 2).

Pose 1 (Figure 5A) exhibits the most crucial ligand–receptor interactions observed for the co-crystallized ligand (Figure 5B). During the MD simulation, Asn253 (6.55) and His278 (7.43) play a major role in the ligand binding process (Figure 5C) and, interestingly, Glu169 is recruited from EL2, forming an additional polar interaction, which has not been evidenced in the corresponding crystallographic structure. It is interesting to compare these data to the results obtained using pose 2 as the starting conformation: Indeed, both poses share an identical orientation of the 1,2,4-triazine scaffold (Figure 5A) into the binding site, with the only difference being the orientation of the chlorophenol moiety, as highlighted in Figure S1 in the Supporting Information. Moreover, both conformations show a very similar average fluctuation profile (RMSD) inside the binding pocket (pose 1: 2.63 ± 0.96 Å; pose 2: 1.78 ± 0.53 Å). Nevertheless, the IEFs reported in Figure S1c in the Supporting Information clearly reveal that the starting orientation of the chlorine substituent is crucial to allow the ligand to establish a stable hydrogen bond interaction with His278 (7.43): In fact, the electrostatic contribution of His278 (7.43) to ligand binding, after 60 ns of MD, is more favorable for pose 1 than for pose 2.

Poses 4 and 5 apparently show strong hydrogen bond interactions with Asn253 (6.55) and His278 (7.43) (see Figure 5B). However, these interactions are not maintained during the MD simulations (Figure 5C): Indeed, the interaction with His278 (7.43) is readily lost and the one involving Asn253 (6.55) weakens progressively. The other conformations show either weaker (poses 8 and 9) or nonexistent (poses 3, 6, 7, and 10) interactions with Asn253 (6.55) and none with His278 (7.43). The only predicted strong interaction is the hydrogen bond between pose 8 and Glu169 (EL2). In all the cases, however, the MD simulations revealed high ligand fluctuations and overall unstable interaction patterns with the only constant interaction being the hydrophobic contact with Phe168 (EL2).

The strong and stable ligand–receptor interaction pattern described for pose 1 is confirmed by the high slope coefficients values of the linear function $f(x) = m \cdot x$, fitted on the DSF data obtained from MD trajectories, reported in Table S3 in the Supporting Information.

These results confirm how difficult it might be to select a proper conformation by taking into account either only the docking score or simply the presence/absence of ligand–receptor contacts. Indeed, our MD simulations have shown that conformations initially sharing similar interaction patterns (e.g.,

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

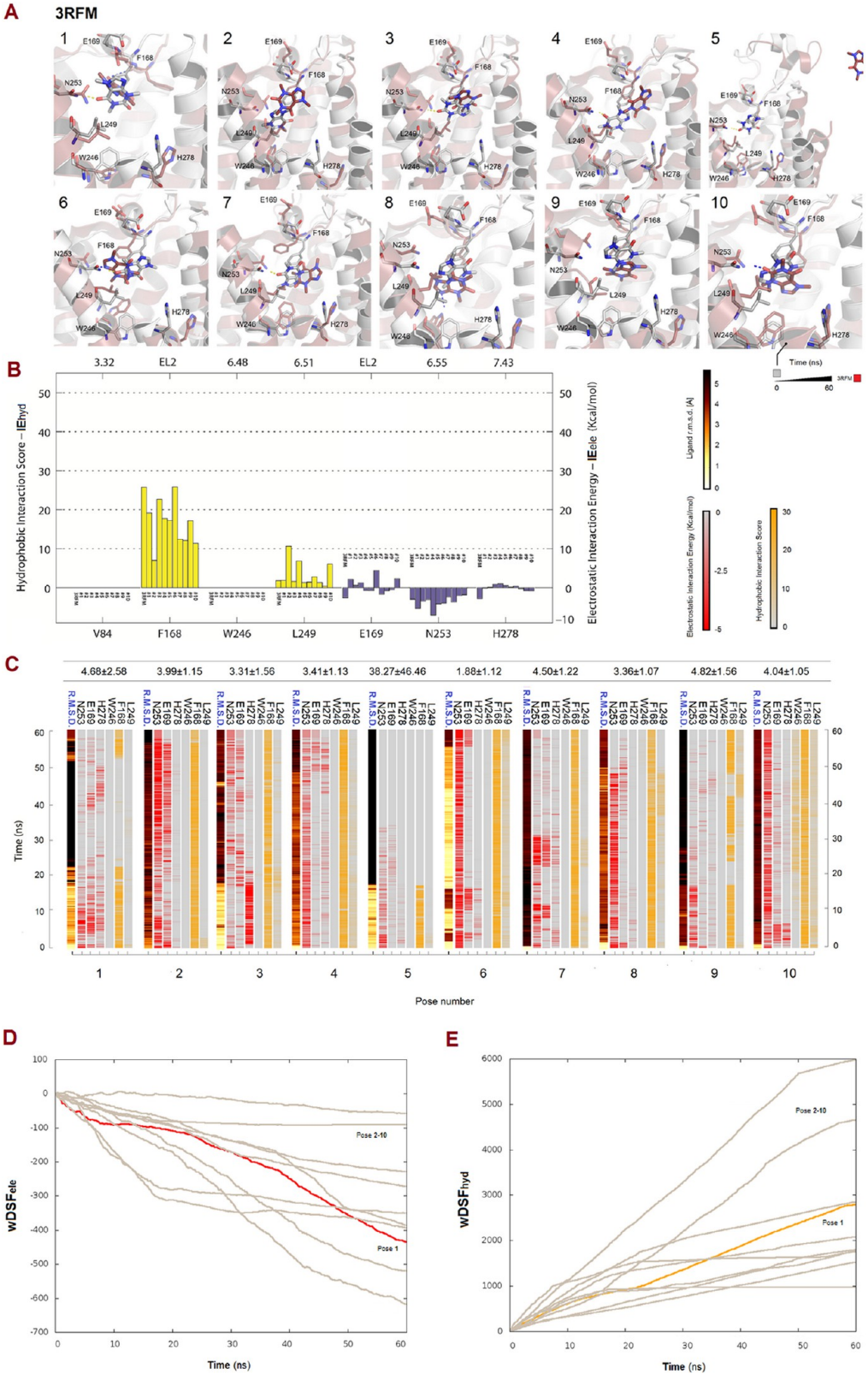


Figure 6. (A) Docking poses of caffeine at the hA_{2A} AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic and (E) hydrophobic wDSFs. IE_{ele} values are given in kcal $Å^{-1}$ mol^{-1}, IE_{hyd} values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

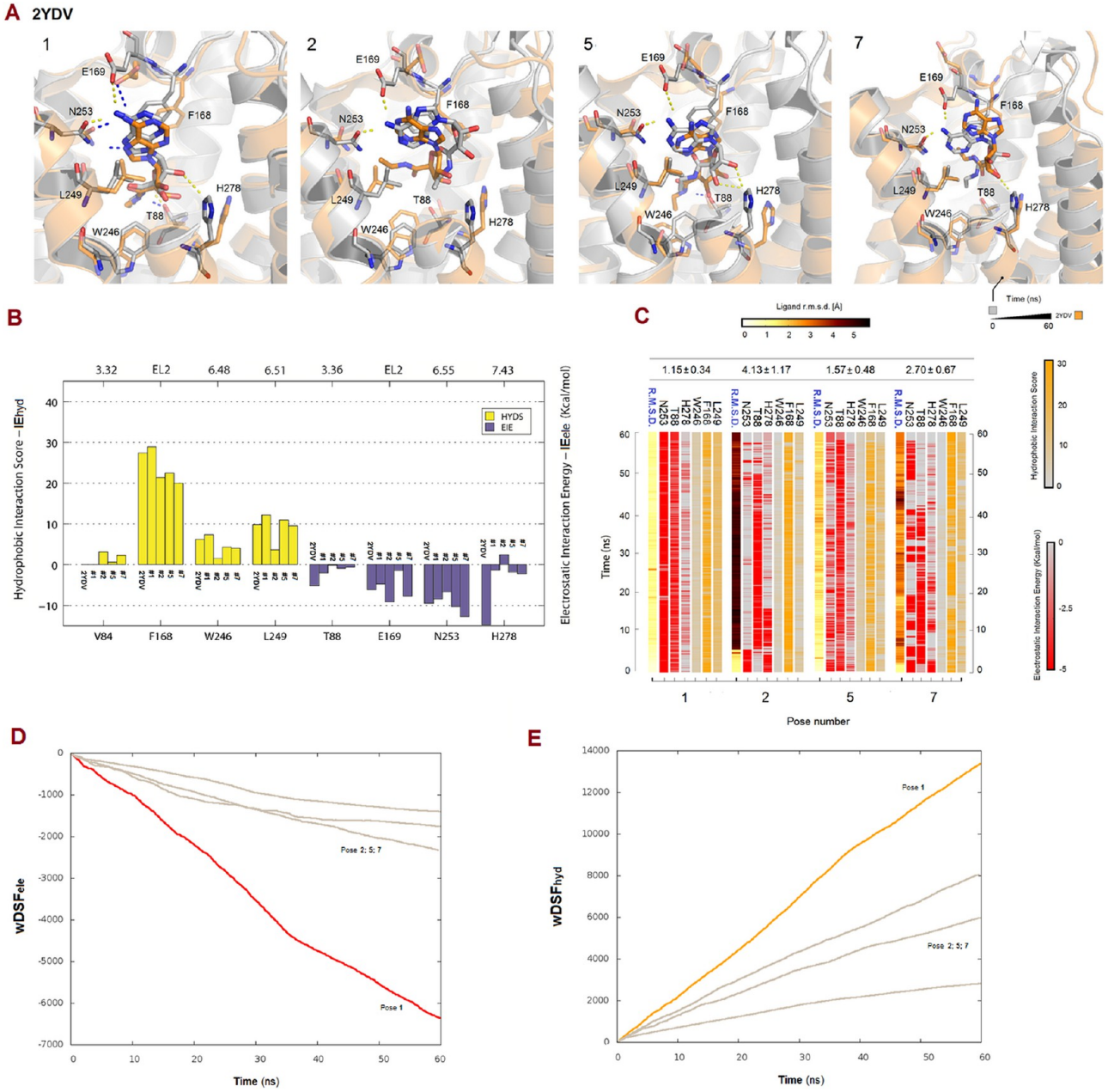


Figure 7. (A) Docking poses of NECA at the hA_{2A} AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic and (E) hydrophobic wDSFs. IE_{ele} values are given in kcal $Å^{-1}$ mol^{-1}, IE_{hyd} values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.

pose 1 and pose 2) may temporally evolve in final states exhibiting different ligand−receptor interactions. This further emphasizes the importance of selecting a proper binding mode, especially in all those cases where no crystal structures are available for comparison (e.g., when the target structure has not yet been solved and one relies on homology models). So, we believe that, in these cases, our proposed protocol might help in selecting the binding mode: In fact, we have recently applied the combined methodology to discern between two possible binding modes of a series of 5-alkylaminopyrazolo[4,3-*e*]1,2,4-triazolo[1,5-*c*]pyrimidine at the hA_3 AR.[53] In that specific case, the MD simulations have helped in supporting a binding mode that was apparently less plausible than the alternative one.

In Silico Inspection of Caffeine Binding Mode. The clustered binding modes of caffeine are very diversified in terms of orientation of the xanthine core, as reported in Figure 6A. However, this variability is not reflected by the calculated docking scores, which do not differentiate the crystallographic conformation (pose 1, RMSD = 0.74 Å) from the others (see Table S1 in the Supporting Information). The calculated IEFs for the retained conformations (Figure 6B) reveal also a lower amount of interactions with key residues, with respect to the aforementioned cases, with the only significant interactions being a hydrogen bond with Asn253 (6.55) and the hydrophobic contacts occurring with the Phe168 (EL2) side chain. This unstable interaction network is further confirmed by

dx.doi.org/10.1021/ci400532b | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

the analysis of the dynamic IEFs (Figure 6C), the ligand positional stability profiles and the collected DSF data (see Figures 6D and 6E and Table S3 in the Supporting Information). These results highlight that it is a difficult task to reproduce the "bioactive" conformation of low affinity ligands showing fragment-like features and lacking strong interactions with the binding site, as, in these cases, there are no energy criteria that can guide the selection of a proper binding mode.

In Silico Inspection of NECA Binding Mode. Four possible binding modes of NECA were retained on the basis of the possible orientations of the adenine ring inside the hA_{2A} AR orthosteric binding pocket (see Figure 7A). Selected poses have RMSD values, with respect to the crystallographic structure, that span from 0.29 Å (pose 1, Table S1 in the Supporting Information) to 4.98 Å (pose 2, Table S1 in the Supporting Information).

Pose 1 (Figure 7A) interacts with residues that hold a crucial role in agonist and antagonist recognition. The side chains of Asn253 (6.55) and Glu169 (EL2) residues establish hydrogen bonds with the exocyclic nitrogen atom of the purine nucleoside derivative. The aromatic purine core, which additionally interacts with Leu249 (6.51), is involved in a $\pi-\pi$ stacking interaction with the conserved Phe168 (EL2) side chain. The ribose moiety, which is deeply inserted in the orthosteric binding pocket, interacts with Thr88 (3.36) and His278 (7.43) through a network of hydrogen bond interactions. During the MD simulation (Figure 7C), the polar contacts with Asn253 (6.55), Glu169 (EL2), and Thr88 (3.36) are maintained strong and persistent, whereas the favorable interaction with His278 (7.43), which is initially involved in a strong interaction with the ligand (Figure 7B), is lost during the simulation. However, the interactions between His278 (7.43) and the ligand are mediated by water molecules, as discussed in the paragraph below.

The dynamic IEFs (Figure 7C) highlight persistent hydrophobic contacts with Leu249 (6.51), Phe168 (EL2) and Trp246 (6.48). The stability of the initial binding mode is also confirmed by the low ligand fluctuation (average RMSD = 1.150 ± 0.34 Å), higher slope coefficients (Table S3 in the Supporting Information) associated with the DSFs (Figures 7D and 7E). The other predicted binding conformations establish hydrogen bond interactions with the Asn253 (6.55) and Glu169 (EL2) side chains (figure 7b). In all these cases, the MD simulations (Figure 7C) have revealed unstable interaction patterns and a lower positional stability, with respect to pose 1, with an increase of the ligand fluctuations into the binding site.

Therefore, despite the fact that different conformations might present similar interaction patterns with residues that play a key role in antagonist or agonist recognition at the hA_{2A} AR (Figure 4B), a "post-processing" selection of the docked poses based on either the docking score or simple visual inspection could lead to misleading results and even failures when building SAR reports. Instead, our proposed analysis of dynamic IEFs, ligand positional stability (RMSD profile), and DSFs has proven to be able to discern the "bioactive" binding mode among other conformations. Indeed, our analysis has highlighted that the conformation with the highest degree of fitness to the hA_{2A} AR binding pocket (which is the one closest to the crystallographic structure) presents a less marked average fluctuation with respect to the other sampled conformations and both stable polar interactions and persistent hydrophobic contacts. The latter aspects could be better understood by analyzing the slope coefficients of the trend line fitted on the DSF data. The conformations that possess an high degree of fitness inside the hA_{2A} AR orthosteric binding pocket are most likely to be characterized by higher slope coefficients (absolute values). These values highlight that strong interactions with crucial residues are maintained through all MD trajectory, thus steadily increasing the cumulative sum of electrostatic interaction energies or hydrophobic score.

In Silico Inspection of Water Molecules Clusters. As previously stated, the scoring functions often fail to properly predict binding affinities, because of their limited description of protein flexibility and the implicit treatment of the solvent. The fully atomistic MD protocol has the advantage to explore along with the hA_{2A} AR−ligand complex flexibility also its dynamical solvation process, thus by providing useful insights into the role of water molecules in the ligand−protein recognition mechanism. The possible structural presence of ordered clusters of water molecules in the proximity of highly conserved motifs in class A GPCRs has been already deeply discussed in the past years[54] and recently clarified for the hA_{2A} AR.[55] Therefore, for each considered ligand−protein complex, we monitored the permanency of water molecules within three different regions, defined as extracellular cluster (EC), central cluster (CC), and intracellular cluster (IC). Here, we describe the EC, whereas descriptions of the CC and IC can be found in the Supporting Information.

The extracellular (EC) cluster (see the left central panel in Figure S3 in the Supporting Information) is located inside the orthosteric binding pocket and plays a crucial role in ligand binding. Hence, we inspected the presence and permanence through the simulation time of unique water molecules within a range of 3.6 Å (donor/acceptor distance) from ligand potential donor or acceptor atoms. In the case of ZM 241385 (PDB ID: 3EML),[15] the presence of a cluster of water molecules plays a role in bridging the ligand to TM2 and TM7 (see Figure S3 in the Supporting Information, upper panel): In particular, TIP161 and TIP6978 (15% permanency) are part of an organized cluster that bridges His278 (7.43) to the nitrogen atom of the triazolotriazine. This might account for the role of His278 (7.43) in the antagonists binding revealed by mutagenesis data,[56] but that has not been yet reported for ZM 241385, although, in recent X-ray structures,[15,55] two water molecules have been observed between the ligand aromatic core and the His278 (7.43) side chain. In the case of NECA (PDB ID: 2YDV[17]), we found the presence of a similar organized cluster of water molecules (TIP1706, TIP3138, and TIP2418) that connects TM7 to the ligand. The dynamic evolution the binding mode of T4G has also highlighted the presence of a water molecule (in rapid exchange) that bridges His278 (7.43) to the nitrogen atom of the 2,6-dimethylpyridin-4-yl substituent (see the upper panel in Figure S3b in the Supporting Information). Water molecules that establish hydrogen bonds with the 1,2,4-triazin-3-amine core have been detected also within a range of 4 Å around T4E, but they are in rapid exchange with other solvation molecules (permanence time <2%). The analysis of the evolution of caffeine binding mode has instead revealed a greater number of water molecules in rapid exchange around the antagonist structure: This is a direct consequence of the weak interactions that the ligand establishes with the protein residues, which make the structure more likely to be surrounded by solvent molecules.

In Silico Inspection of Protein Stability. In addition to the above-described analyses, the overall biophysical stability of

dx.doi.org/10.1021/ci400532b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

the solvated protein−membrane systems has been also assessed. We analyzed in details the conformational stability of each ligand−protein complex by evaluating the fluctuations of the α carbon atoms during the MD simulations. The results reported in Figure S4 in the Supporting Information highlight a common flexibility pattern among all complexes: In particular, transmembrane domains are relatively stable at their starting position (RMSD < 2 Å), whereas intracellular and extracellular loops present a higher flexibility.

CONCLUSIONS

In the present work, we have presented a combined strategy based on the integration of molecular docking and membrane molecular dynamics (MD) simulations. The main aim of our approach has been to merge the rapid sampling of ligand poses into in the binding site—distinctive of docking algorithms—with the thermodynamic accuracy of MD simulations in describing, at the molecular level, the stability of a G protein-coupled receptor (GPCR)−ligand complex embedded into a explicit lipid−water environment.

We selected, as a test case, the human A2A adenosine receptor (hA_{2A} AR) in complex with four antagonists—namely, ZM 241385, T4G, T4E, and caffeine—and one agonist (*N*-ethyl-5′-carboxamido adenosine, NECA), and evaluated the ability of our strategy in reproducing their "bioactive" conformation and in discerning it from other poses generated by the docking protocol. Once a proper conformation has been selected, we evaluated the temporal evolution of the occurring ligand−receptor interactions by introducing the concept of "dynamic IEFs" (where the term "IEFs" represents interaction energy fingerprints).

The above-described results have shown that our post-processing procedure can be regarded as a valuable alternative of conventional scoring functions, as it is able to discern/anticipate the "bioactive" conformation of high affinity ligands and to take into account both the complex flexibility in the membrane environment as well as water-driven interactions, which are two aspects of the binding that docking protocols are not yet able to handle with satisfying accuracy. In addition, our proposed strategy might represent a tool to detect and validate the feasibility of alternative binding conformations, as proposed by the docking algorithm: In this case, indeed, a scoring function-driven selection of the poses might mislead, as highlighted by several examples above-discussed.

We also believe that the proposed strategy can be extended to other GPCRs, as well as to homology models. In the latter case, the selection of a proper binding mode is a difficult task, because of the lack of a reference crystal structure. In such perspective, the analysis of dynamic IEFs and of ligand fluctuation profiles as well as the introduction of a "dynamic scoring function", provided by our combined approach, might represent a valuable help in the choice and represent a valuable tool to generate accurate models of GPCRs in complex with their ligands. In such perspective, we recently applied the herein proposed protocol during the GPCR Structure-Based Homology Modeling and Docking Assessment 2013 (http://gpcr.scripps.edu/GPCRDock2013)[57] and demonstrated how the methodology substantially improves the quality of GPCRs homology models, in terms of ligand−receptor contacts. Therefore, we strongly believe that the proposed protocol might represent an efficient method to improve the quality of homology models for docking and screening applications, with the only crucial requirements being the availability of a high-quality receptor model and a high degree of certainty of residues involved in binding.

ASSOCIATED CONTENT

Supporting Information

Tables summarizing data on the retained docked poses (Table S1), biophysical stability of membrane-embedded ligand-protein systems (Table S2), slope coefficients of linear functions fitted on DSF data (Table S3), supplementary figures (Figures S1−S4), and videos (Videos S1−S5). This material is available free of charge via the Internet at http://pubs.acs.org

AUTHOR INFORMATION

Corresponding Author

*Tel.: +39 049 8275704. Fax: +39 049 827 5366. E-mail: stefano.moro@unipd.it.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The molecular modeling work coordinated by S.M. has been carried out with financial support of the University of Padova, Italy, and the Italian Ministry for University and Research (MIUR), Rome, Italy. S.M. is also very grateful to Chemical Computing Group and Acellera, Ltd., for the scientific and technical partnership.

ABBREVIATIONS

ARs = adenosine receptors; DSF = dynamic scoring function; EL2 = second extracellular loop; GPCRs = G protein-coupled receptors; GPU = graphics processing unit; IEFs = interaction energy fingerprints; MD: molecular dynamics; n.d. = not determined; NECA = *N*-ethyl-5′-carboxamido adenosine; POPC = 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; T4E = 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol; T4G = 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine; TM = transmembrane; ZM 241385 = 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino)-ethyl)phenol

REFERENCES

(1) Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 639−650.
(2) Sodhi, A.; Montaner, S.; Gutkind, J. S. Viral hijacking of G-protein-coupled-receptor signalling networks. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 998−1012.
(3) Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: Molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* **2004**, *103*, 21−80.
(4) Drews, J. Drug discovery: A historical perspective. *Science* **2000**, *287*, 1960−1964.
(5) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727−730.
(6) Jacobson, K. A.; Costanzi, S. New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Mol. Pharmacol.* **2012**, *82*, 361−371.
(7) Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−5931.
(8) Böhm, H.-J.; Stahl, M. The Use of Scoring Functions in Drug Discovery Applications. In *Reviews in Computational Chemistry*;

Lipkowitz, K. B., Boyd, D. B., Eds.; John Wiley & Sons, Inc.: New York, 2003; pp 41−87.
(9) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118−13123.
(10) Dror, R. O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.* **2010**, *135*, 555−562.
(11) Buch, I.; Harvey, M. J.; Giorgino, T.; Anderson, D. P.; De Fabritiis, G. High-throughput all-atom molecular dynamics simulations using distributed computing. *J. Chem. Inf. Model.* **2010**, *50*, 397−403.
(12) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632−1639.
(13) Haskó, G.; Linden, J.; Cronstein, B.; Pacher, P. Adenosine receptors: Therapeutic aspects for inflammatory and immune diseases. *Nat. Rev. Drug Discovery* **2008**, *7*, 759−770.
(14) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **2011**, *19*, 1283−1293.
(15) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211−1217.
(16) Congreve, M.; Andrews, S. P.; Doré, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J. Med. Chem.* **2012**, *55*, 1898−1903.
(17) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, *474*, 521−525.
(18) *Acellera*; http://www.acellera.com/ (accessed Jan 2, 2014).
(19) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relationships in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366−428.
(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(21) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **2010**, *38*, D142−D148.
(22) Jain, E.; Bairoch, A.; Duvaud, S.; Phan, I.; Redaschi, N.; Suzek, B. E.; Martin, M. J.; McGarvey, P.; Gasteiger, E. Infrastructure for the life sciences: Design and implementation of the UniProt website. *BMC Bioinf.* **2009**, *10*, 136.
(23) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.
(24) Labute, P. Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187−205.
(25) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(26) *Molecular Operating Environment*; http://www.chemcomp.com/ (accessed Jan 2, 2014).
(27) Stewart, J. J. P. Optimization of parameters for semiempirical methods. I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.
(28) Stewart, J. J. P. Optimization of parameters for semiempirical methods. II. Applications. *J. Comput. Chem.* **1989**, *10*, 221−264.
(29) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(30) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.
(31) Dowling, J. E.; Vessels, J. T.; Haque, S.; Chang, H. X.; Van Vloten, K.; Kumaravel, G.; Engber, T.; Jin, X.; Phadke, D.; Wang, J.; Ayyub, E.; Petter, R. C. Synthesis of [1,2,4]triazolo[1,5-a]pyrazines as adenosine A2A receptor antagonists. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4809−4813.
(32) Federico, S.; Paoletta, S.; Cheong, S. L.; Pastorin, G.; Cacciari, B.; Stragliotto, S.; Klotz, K. N.; Siegel, J.; Gao, Z. G.; Jacobson, K. A.; Moro, S.; Spalluto, G. Synthesis and Biological Evaluation of a New Series of 1,2,4-Triazolo[1,5-*a*]-1,3,5-triazines as Human A_{2A} Adenosine Receptor Antagonists with Improved Water Solubility. *J. Med. Chem.* **2011**, *54*, 877−89.
(33) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257−265.
(34) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623−625.
(35) Sommer, B. Membrane Packing Problems: A Short Review on Computational Membrane Modeling Methods and Tools. *Comput. Struct. Biotechnol. J.* **2013**, *5*.
(36) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (33−38), 27−28.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(38) Grubmüller, H.; GrollV. *Solvate*; http://www.mpibpc.mpg.de/home/grubmueller/downloads/solvate/index.html (accessed Jan 2, 2014).
(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(40) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52*, 3144−3154.
(41) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52*, 3155−3168.
(42) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503−506.
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision B.01; Gaussian, Inc.: Wallingford, CT, 2009.
(44) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **2001**, *22*, 501−508.

(45) Allen, W. J.; Lemkul, J. A.; Bevan, D. R. GridMAT-MD: A grid-based membrane analysis tool for use with molecular dynamics. *J. Comput. Chem.* **2009**, *30*, 1952−1958.
(46) Kucerka, N.; Tristram-Nagle, S.; Nagle, J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *J. Membr. Biol.* **2005**, *208*, 193−202.
(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(48) Jaakola, V.-P.; Lane, J. R.; Lin, J. Y.; Katritch, V.; Ijzerman, A. P.; Stevens, R. C. Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: Identification and characterization of essential amino acid residues. *J. Biol. Chem.* **2010**, *285*, 13032−13044.
(49) Cristalli, G.; Lambertucci, C.; Marucci, G.; Volpini, R.; Dal Ben, D. A2A adenosine receptor and its modulators: Overview on a druggable GPCR and on structure-activity relationship analysis and binding requirements of agonists and antagonists. *Curr. Pharm. Des.* **2008**, *14*, 1525−1552.
(50) Guvench, O.; MacKerell, A. D., Jr. Computational evaluation of protein-small molecule binding. *Curr. Opin. Struct. Biol.* **2009**, *19*, 56−61.
(51) Jansen, J. M.; Martin, E. J. Target-biased scoring approaches and expert systems in structure-based virtual screening. *Curr. Opin. Chem. Biol.* **2004**, *8*, 359−364.
(52) Wichapong, K.; Lawson, M.; Pianwanit, S.; Kokpol, S.; Sippl, W. Postprocessing of protein-ligand docking poses using linear response MM-PB/SA: application to Wee1 kinase inhibitors. *J. Chem. Inf. Model.* **2010**, *50*, 1574−1588.
(53) Federico, S.; Ciancetta, A.; Sabbadin, D.; Paoletta, S.; Pastorin, G.; Cacciari, B.; Klotz, K. N.; Moro, S.; Spalluto, G. Exploring the Directionality of 5-Substitutions in a New Series of 5-Alkylaminopyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine as a Strategy to Design Novel Human A3 Adenosine Receptors Antagonists. *J. Med. Chem.* **2012**, *55*, 9654−9668.
(54) Pardo, L.; Deupi, X.; Dölker, N.; López-Rodríguez, M. L.; Campillo, M. The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *Chembiochem* **2007**, *8*, 19−24.
(55) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **2012**, *337*, 232−236.
(56) Kim, J.; Wess, J.; Van Rhee, A. M.; Schöneberg, T.; Jacobson, K. A. Site-directed mutagenesis identifies residues involved in ligand recognition in the human A2a adenosine receptor. *J. Biol. Chem.* **1995**, *270*, 13987−13997.
(57) Kufareva, I.; Katrich, V.; GPCR Dock 2013 participants; Stevens, R. C.; Abagyan, R. Advances in GPCR modeling evaluated by the GPCR Dock 2013 Assessment: Meeting new challenges. Submitted to *Structure*.

Original publication III

Sabbadin, D., Moro, *Supervised Molecular Dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale* (2014). Journal of Chemical Information and Modeling In Press.

Supervised Molecular Dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale.

Davide Sabbadin,[a] and Stefano Moro*[a]

[a] *Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università di Padova, via Marzolo 5, 35131 Padova, Italy.*

KEYWORDS: Supervised Molecular Dynamics, Membrane Molecular Dynamics, G Protein-Coupled Receptors; Molecular Docking, Adenosine Receptors.

ABSTRACT: Supervised MD (SuMD) is a computational method that allows the exploration of ligand-receptor recognition pathway investigations in a nanosecond (ns) time scale. It consists of the incorporation of a *tabu-like* supervision algorithm on the ligand-receptor approaching distance into a classic Molecular Dynamics (MD) simulation technique. In addition to speeding up the acquisition of the ligand-receptor trajectory, this implementation facilitates the characterization of multiple binding events (such meta-binding, allosteric and orthosteric sites) by taking advantages of the all-atom MD simulations accuracy of a GPCR-ligand complex embedded into explicit lipid-water environment.

G protein-coupled receptors (GPCRs) are membrane proteins that serve as crucial signal transduction machineries, linking various extracellular inputs with diverse cellular responses. Indeed, a large number of clinically relevant drugs elicit their therapeutic effect(s) through GPCRs[1,2]. During the past few years, crystallography of GPCRs has experienced an unpredictable growth, resulting in the determination of the structures of 20 distinct receptors that, including closely related subtype homology models, this coverage amounts to approximately 12% of the human GPCR superfamily[3]. This high-resolution structural information is helping redefine our knowledge of how GPCRs recognize such a diverse classes of ligands and how they transmit signals across the cell membrane Moreover, they have provided an enormous opportunity for computational methodologies to make major contributions in this field. In particular, molecular dynamics simulations have become a driving factor in many areas of GPCR biophysics and molecular pharmacology, improving our understanding of ligand-receptor interaction, activation mechanisms, receptor hydration and ligand-subtype selectivity[4-8]. Given that computers will continue to get faster and more structures will be solved, the importance of computational methods will only continue to grow, particularly as simulation research is more closely coupled to experiment.

In fact, one of the most challenging issue for the future of drug discovery is the capability to understand the GPCR-ligand recognition pathway with the aim to facilitate the development of drug candidates with more favorable phamacodynamic profiles. Unfortunately, the recognition process between a ligand and its receptor is a very rare event to describe at the molecular level and, even with the recent GPU-based computing resources, it is necessary to carry out classical molecular dynamics (MD) experiments in a long microsecond time scale[5,7]. In order to overcome this limiting factor, we have implemented an alternative MD approach, named Supervised Molecular Dynamics (SuMD) that enables to follow GPCR-ligand approaching process within a time scale reduced, up to 3 orders of magnitude, compared to classical MD. SuMD enable the investigation of ligand-receptor binding events independently from the starting position, chemical structure of the ligand and also from its receptor binding affinity.

SuMD is a standard MD simulation in which the ligand-receptor docking pathway is supervised by a *tabu-like* algorithm (Figure 1). During the production of the MD trajectory the distance between the center of masses of the ligand atoms and the residues composing the orthosteric binding site of the GPCR ($dcm_{L\text{-}R}$) is monitored over a fixed time window (Δt_{ck}, e.g. 200 ps). An arbitrary number of distance points (n: a, b, c, d, e) per each checkpoint trajectory are collected in real time and a linear function $f(x)=m \cdot x$ is fitted on the distance points at the end of the checkpoint time. A supervision *tabu-like* algorithm is applied to increase the probability to produce ligand-receptor binding events without introducing bias to the MD simulation. More precisely, if $m<0$, ligand-receptor distance is likely to be shortened over the checkpoint time, classic MD simulation is restarted from the last produced set of coordinates. Otherwise, the simulation is restored from the original set of coordinates and random velocities of each atom in the system, reassigned coherently to the NVT ensemble. The *tabu-like* supervision algorithm is perpetuated in time until ligand-receptor distance ($dcm_{L\text{-}R}$) is less than 5 Å. To validate the methodology, we selected as a key study the human A_{2A} Adenosine Receptor ($hA_{2A}AR$), that has been recently crystallized with several ligands, both agonists and antagonists, characterized by different receptor binding affinities. In particular, we selected four crystal structures of the $hA_{2A}AR$ in complex with three strong binders such as 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino)ethyl)phenol, ZM 241385 ($pK_D = 9.18 \pm 0.0$[9], PDB code: 3EML[10]); 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G ($pK_D = 8.9 \pm$ n.d.[9], PDB code: 3UZA[11]);

4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pK_D = 9.6 ± n.d.[9], PDB code: 3UZC[11]); and a weaker binder such as caffeine (pK_D = 5.31 ± 0.44[9], PDB code: 3RFM[9]).

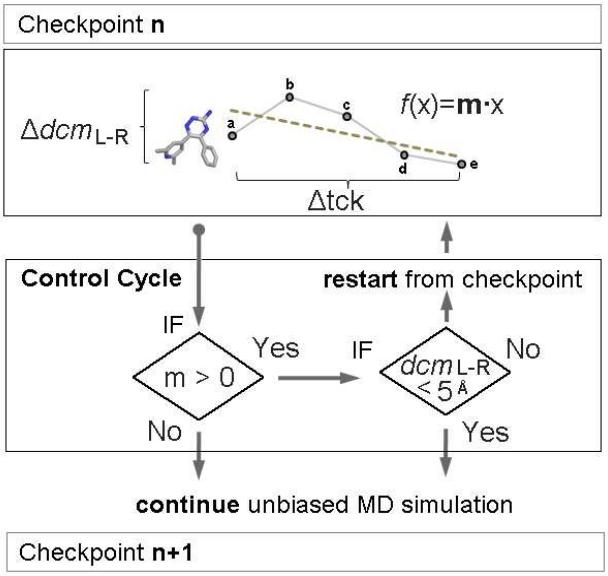


Figure **1**. Scheme of the ligand-receptor distance vector (dcm_{L-R}) supervision algorithm implemented in the Supervised Molecular Dynamics (SuMD) technique.

In all analyzed cases, we were able to reproduce the complete binding process in a nanosecond time scale reproducing with high accuracy the crystallographic pose of each ligand. All SuMD trajectories were run in triplicate (see Supplementary Information for more details). Moreover, using SuMD simulations it is possible to easily determine and characterize all possible ligand binding sites that chronologically anticipate the orthosteric one. These sites are better known as meta-binding sites[12] and in some cases they may coincide with possible allosteric sites.[7] SuMD approach has the potential to facilitate a better understanding of all GPCR-ligand recognition pathway thus increasing the potentiality of *in silico* screening to expedite drug development taking account of full protein flexibility, water-mediated ligand-receptor interactions and the presence of the membrane environment as well.

ZM241385-human A_{2A} Adenosine Receptor recognition mechanism.

Ligand recognition pathway described by Supervised Molecular Dynamics highlight two major interaction sites that anticipate the crystallographic binding conformation (b,c - Figure 2, panel A). In particular Extracellular Loop 2 (EL2) and Extracellular Loop 3 (EL3) of the human A_{2A} Adenosine Receptor are involved in the ligand recognition process. The highlighted meta-binding sites are engaged in tuning ZM241385 orientation and conformation to appropriately reach (d - Figure 2, panel A) and fit (e - Figure 2, panel A) into the orthosteric binding site (Video S1-S2). The antagonist, starting from a randomized set of coordinates at least 40 Å away from protein atoms (a - Figure 2, panel A), reach the orthosteric binding site accurately reproducing the crystallographic pose in less than 60 ns.

The most energetically stable ligand-receptor complex structures (i, ii, iii - Figure 2, panel B and C) were extracted after an energy inspection of the conformational ensemble generated from SuMD simulation after the orthosteric binding site recognition and compared to the XRAY structural information available (Figure 2, panel C). Upon recognition, ZM241385 exibits low fluctuation into the binding site (r.m.s.f. of the triazolotriazine core < 2 Å over 5ns) and the phenolethyl chain, attached to the triazolotriazine ring, explore the same diverse conformational landscape anticipated by XRAY crystallography. In fact, in the latest stage of formation of the high affinity antagonist-human A_{2A} Adenosine Receptor, the structural information extracted from SuMD simulation is undistinguishable (r.m.s.d. below crystallographic resolution limits) from the XRAY crystallographic structure available even in the case of different receptor constructs (Figure 2, panel C).

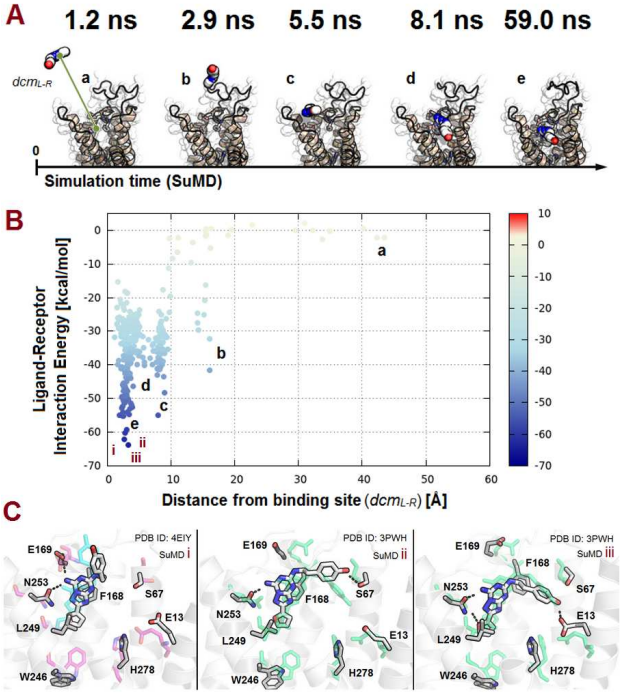


Figure **2.** Panel A: Overview of the Adenosine Receptor Antagonist ZM241385-human A_{2A} Adenosine Receptor recognition mechanism using Supervised Molecular Dynamics (SuMD). Simulation time, when the depicted event occurs, is reported above the ligand-receptor representation. Ligand-Receptor distance vector (dcm_{L-R}) is shown. Van der Waals spheres represent ZM241385 atoms and receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). Hydrogen atoms are not displayed. Panel B: Ligand-receptor interaction energy landscape for the non-biased ZM241385-human A_{2A} Adenosine Receptor recognition process. Some of the most important characterized binding sites (b, c, d) that anticipate the crystallographic information (e) are highlighted. Interaction energy values are expressed in kcal/mol. Panel C: Overview of the three most energetically stable binding conformation of ZM241385 inside the hA_{2A} AR binding pocket generated from SuMD simulation (white sticks) in comparison with two representative XRAY structures, PDB ID:3PWH (green sticks) and PDB ID:4EIY (cyan

sticks). The complexes are viewed from the membrane side facing TM6 and TM7 with the view of TM7 is partially omitted. Hydrogen atoms are not displayed, whereas hydrogen bond interactions are highlighted as yellow dashed lines.

T4G-human A_{2A} Adenosine Receptor recognition mechanism.

As for ZM241385, T4G recognition pathway highlights multiple ligand-receptor binding events that anticipate the orthosteric binding site recognition. The 1,2,4-triazine derivative can be trapped in a transient pocket, named meta-binding site 2 (Video S3,S4), located in the Second Extracellular Loop 2 (ECL2). After this binding event (b – Figure 3, panel A), the aromatic substituents at the 1,2,4-triazine aromatic core are directed towards the third extracellular loop (EL3) which represent the common meta-binding site (c – Figure 3, panel A) explored by ZM241385 as well. In this case the antagonist, starting from a randomized set of coordinates at least 40 Å away from protein atoms (a – Figure 3, panel A), reach the orthosteric binding site (d – Figure 3, panel A) accurately reproducing the crystallographic pose (e – Figure 3, panel A) in less than 65 ns (see Supplementary Information for more details).

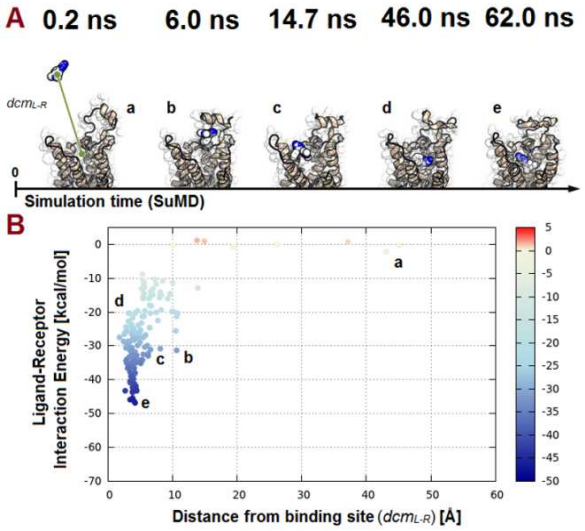


Figure **3**. Panel A: T4G-human A_{2A} Adenosine Receptor recognition mechanism using Supervised Molecular Dynamics (SuMD). Simulation time, when the depicted event occurs, is reported above the ligand-receptor representation. Ligand-Receptor distance vector (dcm_{L-R}) is shown. Van der Waals spheres represent T4G atoms and receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). Hydrogen atoms are not displayed. Panel B: Ligand-receptor interaction energy landscape for the non-biased T4G-human A_{2A} Adenosine Receptor recognition process. Some of the most important characterized binding sites (b, c) that anticipate the crystallographic information (e) are highlighted. Interaction energy values are in kcal/mol.

T4E-human A_{2A} Adenosine Receptor recognition mechanism.

The Extracellular Loop 3 (EL3) of the human A_{2A} Adenosine Receptor play a crucial role in the recognition of T4E. In fact, starting from a randomized set of coordinates at least 40 Å away from protein atoms (a – Figure 4, panel A), the antagonist make contact with EL3 (b, c – Figure 4, panel A) and eventually reach the orthosteric binding site (d – Figure 4, panel A) and make contacts that accurately reproduce the crystallographic structure (e – Figure 4, panel A) in less than 110 ns. The recognition mechanism, investigated using SuMD, is reported in the supplementary information videos S5 and S6.

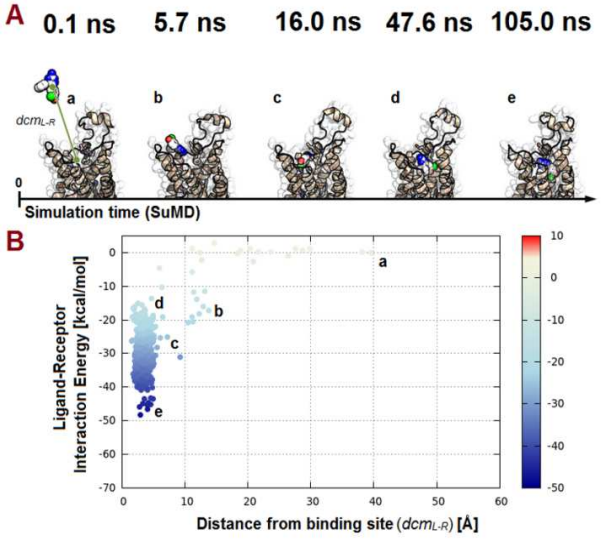


Figure **4**. Panel A. T4E-human A_{2A} Adenosine Receptor recognition mechanism using Supervised Molecular Dynamics (SuMD). Simulation time, when the depicted event occurs, is reported above the ligand-receptor representation. Ligand-Receptor distance vector (dcm_{L-R}) is shown. Van der Waals spheres represent T4E atoms and receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). Hydrogen atoms are not displayed. Panel B. Ligand-receptor interaction energy landscape for the non-biased T4E-human A_{2A} Adenosine Receptor recognition process. Some of the most important characterized binding sites (b, c) that anticipate the crystallographic information (e) are highlighted. Interaction energy values are in kcal/mol.

Caffeine-human A_{2A} Adenosine Receptor recognition mechanism.

As reported for the high-affinity human A_{2A} Adenosine Receptor antagonists, the purine derivative Caffeine recognition mechanism is mediated (b – Figure 5, panel A) by the Extracellular Loop 3 (EL3). Upon binding (c, d – Figure 5, panel A) the weak antagonist shows fragment-like features and lack strong interactions with the binding site (rmsf. > 4 Å). The complete binding event, described using SuMD, is reported in the supplementary information videos S7 and S8.

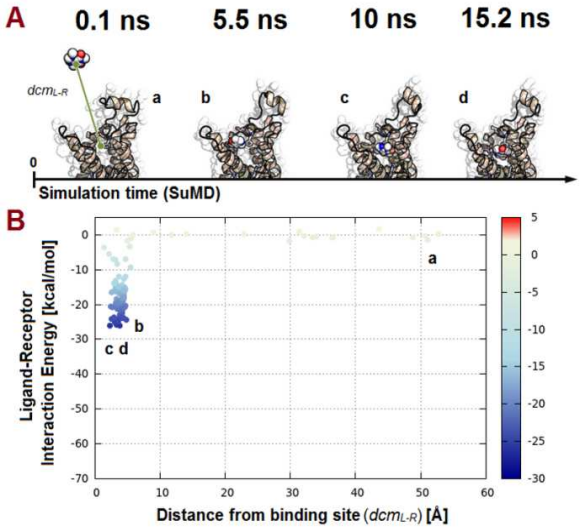


Figure **5**. Panel A: Caffeine-human A_{2A} Adenosine Receptor recognition mechanism using Supervised Molecular Dynamics (SuMD). Ligand-Receptor distance vector (dcm_{L-R}) is shown. Van der Waals spheres represent caffeine atoms and receptor ribbon representation is viewed from the membrane side facing transmembrane domain 6 (TM6) and transmembrane domain 7 (TM7). Hydrogen atoms are not displayed. Panel B: Ligand-receptor interaction energy landscape for the non-biased caffeine-human A_{2A} Adenosine Receptor recognition process. Some of the most important characterized binding sites (b, c) that anticipate the crystallographic information (e) are highlighted. Interaction energy values are in kcal/mol.

Moreover, supervised Molecular Dynamics simulations recognize the critical role of the hA2A AR extracellular loops in the ligand recognition process that have been postulated, using site-directed mutagenesis, in the past[12-14]. The complex evolving network of interactions has been depicted using a simplified ribbon representation of the receptor that comprise a quantitative estimate of the occurring ligand-protein mutual recognition process (Figure 6 and 7). In fact, SuMD could represent a powerful tool to assist the design a focused set of aminoacid mutation experiments in order to infer their role on the molecular recognition process. A critical analysis of the interaction maps reported in detail in Figure 6 and 7, highlight the involvement of the vast majority of the residues located in the Extracellular Loop 2 (EL2) and Extracellular Loop 3 (EL3) of hA2A AR in ligand recognition, thus confirming the crucial role of the acidic residues located in EL2[13] (E151, D170, E169).

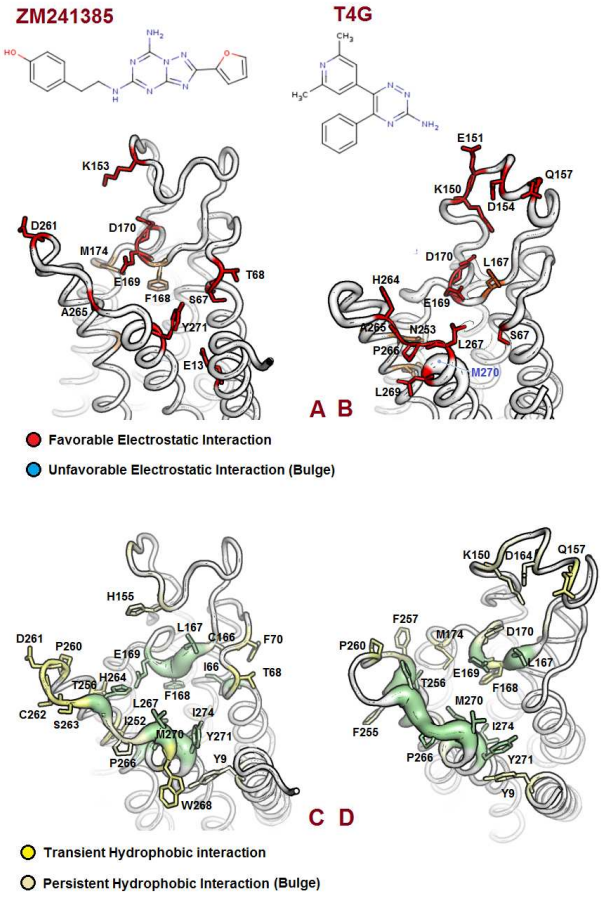


Figure **6.** Electrostatic (panel A and B) and hydrophobic (panel C and D) contributions to the interaction energy of each receptor residue, involved in the binding with the high affinity hA2A AR antagonists ZM241385 and T4G, during the metabinding sites recognition process. Contributions to ligand binding were calculated during the first 15ns of SuMD simulations. Ribbon representation is viewed from the extracellular side and hydrogen atoms are not displayed.

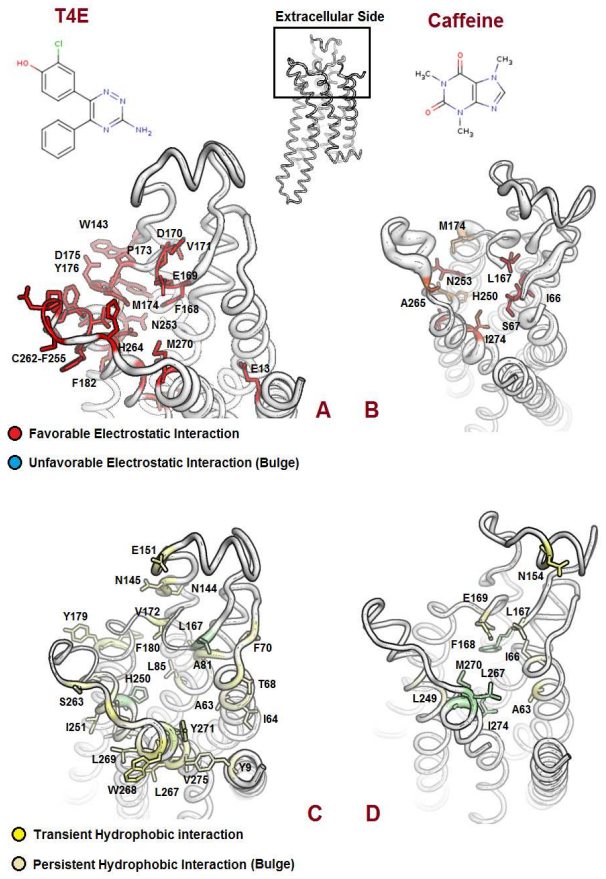


Figure **7.** Electrostatic (panel A and B) and hydrophobic (panel C and D) contributions to the interaction energy of each receptor residue, involved in the binding with the high affinity hA2A AR antagonist T4E and low affinity caffeine, during the meta-binding sites recognition process. Contributions to ligand binding were calculated during the first 15ns of SuMD simulations (10 ns for caffeine). Ribbon representation is viewed from the extracellular side and hydrogen atoms are not displayed.

Concluding, in the present work we have presented a helpful approach to simulate a receptor-ligand recognition pathway based on GPU-driven membrane molecular dynamics simulation in an appreciable short time scale. Combining two very well known and consolidated computational approaches such as molecular dynamics and an a tabu-like algorithm to supervise the evolution of receptor-ligand trajectory, we are able to accurately completely explore the receptor-ligand event in a nanosecond time scale. This approach is also very useful to analyze both orthosteric and allosteric binding events broadening our perspectives in several scientific areas from molecular pharmacology to drug discovery. In particular Supervised MD (SuMD) can be applied in a drug design campaign, for lead optimization, in a high.throughput level in order to design novel binders with preferable pharmacodynamic profiles. Moreover, SuMD represent a powerful tool to assist the design site-directed mutagenesis experiments in order to investigate the molecular recognition process.

ASSOCIATED CONTENT

Supporting Information. Complete experimental section and additional result discussion for ZM241385, T4E, T4G and caffeine- human A_{2A} Adenosine Receptor recognition mechanism is available on the supplementary information material. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università di Padova, via Marzolo 5, 35131 Padova, Italy. Fax: +39 049 8275366; Tel: +39 049 8275704; E-mail: stefano.moro@unipd.it

Funding Sources

This work was supported by University of Padova, Italy, and the Italian Ministry for University and Research (MIUR), Rome, Italy.

ABBREVIATIONS

ARs: Adenosine Receptors; EL2: Second Extracellular Loop; EL3: Third Extracellular Loop; GPCRs: G Protein-Coupled Receptors; GPU: graphics processing unit; MD: molecular dynamics; SuMD: supervised molecular dynamics ; n.d.: not determined; POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; T4E: 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol; T4G: 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine; TM: transmembrane; ZM 241385: 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-a)(1,3,5)triazin-5-yl-amino)ethyl)phenol.

REFERENCES

(1) Hopkins, A. L.; Groom, C. R. The Druggable Genome. *Nat Rev Drug Discov* **2002**, *1*, 727–730.

(2) Drews, J. Drug Discovery: a Historical Perspective. *Science* **2000**, *287*, 1960–1964.

(3) Jacobson, K. A.; Costanzi, S. New Insights for Drug Design from the X-ray Crystallographic Structures of G-protein-coupled Receptors. *Mol. Pharmacol.* **2012**, *82*, 361–371.

(4) Dror, R. O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring Atomic Resolution Physiology on a Femtosecond to Millisecond Timescale Using Molecular Dynamics Simulations. *J. Gen. Physiol.* **2010**, *135*, 555–562.

(5) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and Mechanism of Drug Binding to G-protein-coupled Receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118–13123.

(6) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.

(7) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J.R.; Pan, A.C.; Arlow, D. H.; Canals, M.; Lane, J.R.; Rahmani, R.; Baell, J. B.; Sexton, P. M.; Christopoulos, A.; Shaw, D. E. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **2013**, *503*, 295-299.

(8) Selvam, B.; Wereszczynski, J.; Tikhonova, I. G. Comparison of Dynamics of Extracellular Accesses to the β1 and β2 Adrenoceptors Binding Sites Uncovers the Potential of Kinetic Basis of Antagonist Selectivity. *Chemical Biology & Drug Design* **2012**, *80*, 215–226.

(9) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F. et al. Structure of the Adenosine A(2A) Receptor in Complex with ZM241385 and the Xanthines XAC and Caffeine. *Structure* **2011**, *19*, 1283–1293.

(10) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **2008**, *322*, 1211–1217.

(11) Congreve, M.; Andrews, S. P.; Doré, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A. et al. Discovery of 1,2,4-triazine Derivatives as Adenosine A(2A) Antagonists Using Structure Based Drug Design. *J. Med. Chem.* **2012**, *55*, 1898–1903.

(12) Moro, S.; Hoffmann, C.; Jacobson, K. A. Role of the extracellular loops of G protein-coupled receptors in ligand recognition: a molecular modeling study of the human P2Y1 receptor. *Biochemistry* **1999**, *38*, 3498-3507.

(13) Kim, J.; Jiang, Q.; Glashofer, M.; Yehle, S.; Wess, J.; Jacobson, K. A. Glutamate Residues in the Second Extracellular Loop of the Human A2A Adenosine Receptor Are Required for Ligand Recognition. *Mol. Pharmacol.* **1996**, *49*, 683–691.

(14) Kim, J.; Wess, J.; van Rhee, A. M.; Schöneberg, T.; Jacobson, K. A. Site-Directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A2a Adenosine Receptor. *J. Biol. Chem.* **1995**, *270*, 13987–13997.

3 Conclusion and future perspectives

In this thesis work we propose different integrated approaches that can have potential breakthrough impact across a wide variety of application into the Molecular Modeling field. We expect that the XRAY crystal structures available for GPCRs will be dramatically increasing over the next years. Adenosiland platform (http://mms.dsfarm.unipd.it/Adenosiland) concept, being the first integrated bioinformatics and chemoinformatics web-resource dedicated to adenosine receptors will be extended to all the GPCRs family. The possibility to analyze receptor-membrane systems will be extended to all GPCRs of which structural information will become available. Users will have access to explore structural diversity from an evolutionary point of view. Similarity search screenings against all GPCRs agonists and antagonists deposited in ChEMBLdb with co-crystallized ligands will be delivered on the Adenosiland 2.0 version. We believe that Adenosiland platform will be a starting point for the beginning of new web-platforms that provide experimental data and computational predictions thus being valuable information for the rational design of desired highly potent and selective ligands.
Moreover the presented methodology that enables to overcome scoring function limitation, reproducing the bio-active binding conformation from an ensemble of structural decoys, take advantage of state-of-the-art technology and could represent a tool of crucial importance in medicinal chemistry research. In such perspective, the proposed methodological advances have been used in the the GPCR Structure-Based Homology Modeling and Docking Assessment 2013 (http://gpcr. scripps.edu/GPCRDock2013) and proved substantial improvement in comparison to the known techniques. GPU-accelerated Molecular Dynamics resulted to represent an efficient method to improve the quality of homology models for docking and screening applications.
Supervised Molecular Dynamics, that take advantage of the full potential of GPU-accelerated Molecular Dynamics, allow the characterization of meta-binding sites and can help the characterization of the ligand-receptor binding-pathway in less than two order of magnitude compared to classical MD simulations. This promising technique can have potential development and implementation on molecular modeling programs that are widely used in both industry and academia.

Appendix 1: Implementation of the Best Template Searching tool into Adenosiland platform

Floris, M., **Sabbadin, D.**, Ciancetta, A., Medda, R., Cuzzolin, A., Moro, S. *Implementing the Best Template Searching tool into Adenosiland platform* (2014) In Silico Pharmacology. In press.

Floris et al. In Silico Pharmacology 2013, 1:25
http://www.in-silico-pharmacology.com/content/1/1/25

In Silico Pharmacology
a SpringerOpen Journal

SHORT REPORT Open Access

Implementing the "Best Template Searching" tool into Adenosiland platform

Matteo Floris[1], Davide Sabbadin[2], Antonella Ciancetta[2], Ricardo Medda[1], Alberto Cuzzolin[2] and Stefano Moro[2*]

Abstract

Background: Adenosine receptors (ARs) belong to the G protein-coupled receptors (GCPRs) family. The recent release of X-ray structures of the human A_{2A} AR (h A_{2A} AR) in complex with agonists and antagonists has increased the application of structure-based drug design approaches to this class of receptors. Among them, homology modeling represents the method of choice to gather structural information on the other receptor subtypes, namely A_1, A_{2B}, and A_3 ARs. With the aim of helping users in the selection of either a template to build its own models or ARs homology models publicly available on our platform, we implemented our web-resource dedicated to ARs, *Adenosiland*, with the "*Best Template Searching*" facility. This tool is freely accessible at the following web address: http://mms.dsfarm.unipd.it/Adenosiland/ligand.php.

Findings: The template suggestions and homology models provided by the "*Best Template Searching*" tool are guided by the similarity of a query structure (putative or known ARs ligand) with all ligands co-crystallized with hA_{2A} AR subtype. The tool computes several similarity indexes and sort the outcoming results according to the index selected by the user.

Conclusions: We have implemented our web-resource dedicated to ARs *Adenosiland* with the "*Best Template Searching*" facility, a tool to guide template and models selection for hARs modelling. The underlying idea of our new facility, that is the selection of a template (or models built upon a template) whose co-crystallized ligand shares the highest similarity with the query structure, can be easily extended to other GPCRs.

Keywords: G protein-coupled receptors; Adenosine receptors; Receptor modelling; Bioinformatics platform; Adenosiland

Findings

The template suggestions and homology models provided by the "*Best Template Searching*"tool are guided by the similarity of a query structure (putative or known ARs ligand) with all ligands co-crystallized with hA_{2A} AR subtype. The tool computes several similarity indexes and sort the outcoming results according to the index selected by the user.

Background

Adenosine receptors (ARs) belong to the G protein-coupled receptors (GCPRs) family. The known four subtypes, termed adenosine A_1, A_{2A}, A_{2B} and A_3 receptors, are widely distributed in human body and involved in several physio-pathological processes (Fredholm et al. 2001). The release of X-ray structures of the human A_{2A} AR in complex with agonists (Lebon et al. 2011, Xu et al. 2011) and antagonists (Jaakola et al. 2008, Doré et al. 2011, Hino et al. 2012, Congreve, et al. 2012, Liu, et al. 2012) has enabled to extend structure-based drug design approaches to this class of receptors. With the use of homology modelling techniques, indeed, structural information on the other subtypes can also be derived. As a key step when building homology models is the selection of a proper template, we have developed a tool to guide the user in this crucial choice by implementing the "*Best Template Searching*" facility in our web-resource dedicated to ARs, *Adenosiland* (Floris et al. 2013). This tool is freely accessible at the following web address: http://mms.dsfarm.unipd.it/Adenosiland/ligand.php.

The underlying idea behind this facility is to help the user in selecting the best template or ARs model to get the highest quality receptor for further molecular

* Correspondence: stefano.moro@unipd.it
[2]Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, via Marzolo 5, I-35131 Padova, Italy
Full list of author information is available at the end of the article

Table 1 Values of the in-house validation of the combined similarity index

Input ligand	Suggested template	Combined similarity value
Adenosine	2YDO	0.83
NECA	2YDO	0.72
UK-432,097	3QAK	0.37
ZMA 241385	4EIY	0.69
T4G	3UZA	0.84
T4E	3UZC	0.92
XAC	3REY	0.67
Caffeine	3RFM	0.98

docking studies. A possible strategy herein presented is to compute the similarity between a known or putative agonist/antagonist and all co-crystallized ARs ligands.

Tool description

The "*Best Template Searching*" tool works as follows: the user is asked to input a query molecule either by uploading a SMILES string or by directly drawing the 2D structure by using the JME interface; the similarity of the input molecule is then computed against all the ligands co-crystallized with the hA_{2A} AR. The following similarity indexes are calculated: *(i)* shape similarity (based on the Manhattan distance between USR descriptors), *(ii)* 2D similarity (based on the Tanimoto and Tversky Similarities of Pubchem Fingerprints), *(iii)* pharmacophoric similarity (based on the Tanimoto similarity of Pharmacophoric triplets), and *(iv)* a combined similarity (derived by the following function: 0.6 * pharmacophoric similarity + 0.4 * shape similarity).

The values of the two coefficients composing the latter similarity index have been derived by running a preliminary in-house validation based on all available crystallographic structures: In particular, the two values have been chosen so that by providing as input the structures of the co-crystallized ligand the corresponding receptor structure results the best ranked one according to the combined similarity index. The values obtained for the structures considered for the internal validation are reported in Table 1. For all the structure except one, the suggested template results the corresponding crystal structure. The only exception is represented by NECA for which the structure co-crystallized with adenosine is suggested as best template. Considering the high structural similarity between the two agonist structures, the results is in line with the others.

Simultaneously to the best template searching process, a similarity search screening is also performed against all adenosine agonists and antagonists deposited in ChEMBL, release 14 (Gaulton et al. 2011). In more details, the query is compared to 760 A_1, 469 A_{2A}, 559 A_{2B} and 290 A_3 AR ligands and the comparison is based on the calculation of the similarity measures previously described. The identified compounds are reported in a table along with the associated binding data available in literature.

Tool validation

Ligand similarity biased template selection criteria at the basis of the *"Best Template Searching"* tool has been successfully applied to rationalize the Structure Activity Relationships (SAR) of a series of [5-substituted-4-phenyl-1,3-thiazol-2-yl] furamides as antagonist of the hARs (Inamdar et al. 2013). The most potent derivative of the furamides series, the furan-2-carboxylic acid (4-phenyl-5-pyridin-4-yl-thiazol-2-yl)-amide, has been selected as query molecule: As reported in Table 2, a similarity sorting of the templates based on the combined similarity criteria has been taken into account to select the most suitable models for receptor-based ligand design. The selected workflow is summarized in Figure 1: Starting from the

Table 2 Similarity sorting of human A_{2A} AR templates based on furan-2-carboxylic acid (4-phenyl-5-pyridin-4-yl-thiazol-2-yl)-amide query ligand

Ligand	PDB ID template	Shape similarity	2D similarity (Tanimoto)	2D similarity (Tversky)	Pharmacophore similarity (Tanimoto)	Pharmacophore similarity (Tversky)	Combined similarity (Shape & FP)
T4G	**3UZA**	**0.33**	**0.86**	**0.89**	**0.46**	**0.65**	**0.52**
ZM 241385	3PWH	0.58	0.90	0.93	0.27	0.42	0.48
T4E	3UZC	0.37	0.84	0.89	0.44	0.54	0.47
ZM 241385	4EIY	0.34	0.90	0.93	0.27	0.43	0.39
ZM 241385	3EML	0.35	0.90	0.93	0.27	0.42	0.39
NECA	2YDV	0.51	0.82	0.87	0.17	0.31	0.39
ZM 241385	3VG9	0.32	0.90	0.93	0.27	0.43	0.38
XAC	3REY	0.21	0.89	0.94	0.25	0.48	0.37
ZM 241385	3VGA	0.28	0.90	0.93	0.27	0.42	0.36
Adenosine	2YDO	0.33	0.82	0.86	0.18	0.31	0.31
Caffeine	3RFM	0.26	0.81	0.85	0.21	0.34	0.30
UK-432,097	3QAK	0.16	0.87	0.93	0.14	0.35	0.27

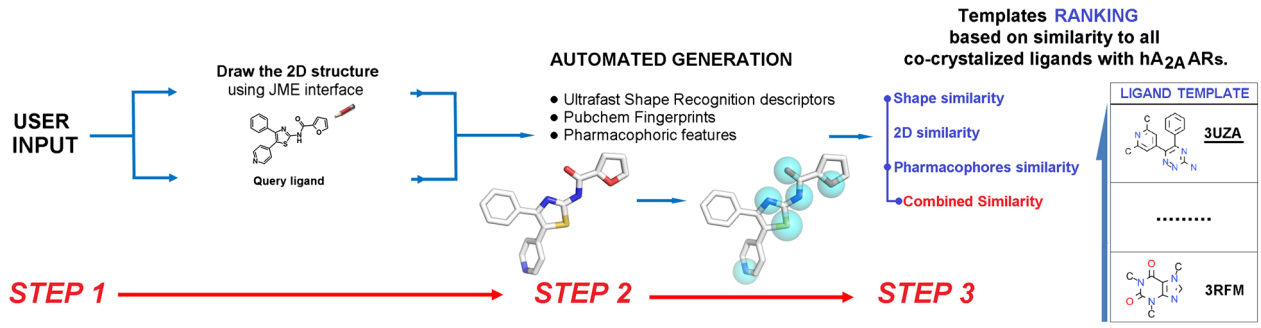


Figure 1 Workflow of the homology modeling template selection based on the structure of furan-2-carboxylic acid (4-phenyl-5-pyridin-4-yl-thiazol-2-yl)-amide.

suggested best template, namely the structure with the 3UZA PDB ID, co-crystallized with the 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (T4G), we have constructed A_1, A_{2B} and A_3 AR models through homology modeling and used the so derived structural information to provide hypotheses of ligand-receptor interaction and ligand-receptor selectivity profile (Inamdar et al. 2013).

Methods

The *"Best Template Searching"* tool is part of the Adenosiland infrastructure, based on Ubuntu 9.10 Linux operating system, which is a patchwork of several informatics tools (for more details see Floris et al. 2013). The similarity indexes are calculated by using different approaches: 2D similarity based on Tanimoto and Tversky indexes (Steinbeck et al. 2003, 2006) are calculated from Pubchem Fingerprints (CDK implementation), the shape similarity is calculated by using an in-house implementation of the Ultrafast Shape Recognition method (Floris et al. 2011, Ballester and Richards 2007), and the pharmacophoric features of the pharmacophore-based similarity index are described by Gaussian 3D volumes (Taminau et al. 2008).

Conclusions

We have implemented a novel tool, called *"Best Template Searching"* to provide template suggestions and homology models of all four hARs based on the similarity between a query structure provided by the user and all co-crystallized ARs ligands. It is well known that ligand-driven induced fit of the receptor is a key feature to facilitate the identification or the optimization of novel potent and selective agonists and antagonists, in particular through molecular docking studies. We therefore believe that choosing as template the structure co-crystallized with the ligand that shares the highest structural similarity with the scaffold of interest may represent an effective strategy. This is in facts the underlying idea of our platform implementation: By using the *"Best Template Searching"* option, users can upload a SMILES string or directly draw the 2D structure by using the JME interface of the scaffold of interest and search the most similar ligand co-crystallized so far with the hA_{2A} AR. Several similarity indexes are calculated by using different approaches such as a 2D similarity, shape similarity, pharmacophore-based similarity, and simple consensus shape- and pharmacophore-based similarity index.

We are also confident that the proposed strategy can be easily and effectively extended to other GPCRs.

Abbreviations

ARs: Adenosine receptors; GPCRs: G protein-coupled receptors; NECA: N-ethyl-5′-carboxamido adenosine; T4E: 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol; T4G: 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine; ZM 241385: 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino) ethyl)phenol; XAC: N-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl- 2,3,6,7-tetrahydro-1H-purin-8-yl)phenoxy]acetamide.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MF, DS and RM developed and engineered the web tool. DS, ACi and ACu carried out the experiments, analyzed the data, and interpreted the results. ACi and SM designed the research protocol and wrote the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

The molecular modeling work coordinated by S.M. has been carried out with financial support of the University of Padova, Italy and the Italian Ministry for University and Research (MIUR), Rome, Italy. S.M. is also very grateful to Chemical Computing Group, YASARA Biosciences GmbH and Acellera for the scientific and technical partnership. Finally, we desire to give our appreciations to Peter Ertl for his courtesy in using its the Java Molecule Editor (JME).

Author details

[1]CRS4, Parco Polaris, 09010 Pula, CA, Italy. [2]Molecular Modeling Section (MMS), Department of Pharmaceutical and Pharmacological Sciences, University of Padova, via Marzolo 5, I-35131 Padova, Italy.

Received: 31 October 2013 Accepted: 25 November 2013
Published: 20 December 2013

References

Ballester PJ, Richards WG (2007) Ultrafast shape recognition to search compound databases for similar molecular shapes. J Comput Chem 28:1711–1723

Congreve M, Andrews SP, Doré AS, Hollenstein K, Hurrell E, Langmead CJ, Mason JS, Ng IW, Tehan B, Zhukov A, Weir M, Marshall FH (2012) Discovery of 1, 2,4-triazine derivatives as adenosine A_{2A} antagonists using structure based drug design. J Med Chem 55:1898–1903

Doré AS, Robertson N, Errey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K, Congreve M, Magnani F, Tate CG, Weir M, Marshall FH (2011) Structure of the adenosine A_{2A} receptor in complex with ZM241385 and the xanthines XAC and caffeine. Structure 19:1283–1293

Floris M, Masciocchi J, Fanton M, Moro S (2011) Swimming into peptidomimetic chemical space using pepMMsMIMIC. Nucleic Acids Res 39:W261–W269

Floris M, Sabbadin D, Medda R, Bulfone A, Moro S (2013) Adenosiland: walking through adenosine receptors landscape. Eur J Med Chem 58:248–257

Fredholm BB, IJzerman AP, Jacobson KA, Klotz KN, Linden J (2001) International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. Pharmacol Rev 53:527–552

Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2011) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107

Hino T, Arakawa T, Iwanari H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y, Kusano-Arai O, Weyand S, Shimamura T, Nomura N, Cameron AD, Kobayashi T, Hamakubo T, Iwata S, Murata T (2012) G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. Nature 482:237–240

Inamdar GS, Pandya AN, Thakar HM, Sudarsanam V, Kachler S, Sabbadin D, Moro S, Klotz K-N, Vasu KK (2013) New insight into adenosine receptors selectivity derived from a novel series of [5-substituted-4-phenyl-1,3-thiazol-2-yl] benzamides and furamides. Eur J Med Chem 63:924–934

Jaakola VP, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC (2008) The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to antagonist. Science 322:1211–1217

Lebon G, Warne T, Edwards PC, Bennett K, Langmead CJ, Leslie AGW, Tate CG (2011) Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation. Nature 474:521–525

Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB, Heitman LH, IJzerman AP, Cherezov V, Stevens RC (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. Science 337:232–236

Steinbeck C, Han Y, Kuhn S, Horlacher O, Luttmann E, Willighagen E (2003) The chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics. J Chem Inf Comput Sci 43:493–500

Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL (2006) Recent developments of the chemistry development kit (CDK) an open-source java library for chemo- and bioinformatics. Curr Pharm Des 12:2111–2120

Taminau J, Thijs G, De Winter H (2008) Pharao: pharmacophore alignment and Optimization. J Mol Graph Model 27:161–169

Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao Z-G, Cherezov V, Stevens RC (2011) Structure of an agonist-bound human A_{2A} adenosine receptor. Science 332:322–327

doi:10.1186/2193-9616-1-25
Cite this article as: Floris *et al.*: **Implementing the "Best Template Searching" tool into Adenosiland platform.** *In Silico Pharmacology* 2013 **1**:25.

Appendix 2: GPCRs dynamic solvation process: insights using all-atom MD simulations

Sabbadin, D., Moro, S. *Hydrodynamic 2D/3D-mapping of protein solvation profile using GPU-driven all-atoms Molecular Dynamics* (2013). Journal of Chemical Information and Modeling (Submitted).

Author's note: Water is the major component of living cells play a crucial role in protein architecture, dynamics and ligand recognition [67]. In absence of a bound ligand, the binding site of a receptor is usually occupied by water molecules that can be displaced upon binding. The energetic cost related to desolvation plays an important role in the design of G-Protein coupled receptors targeting ligand [68]. However, in absence of an high-resolution XRAY structure it is difficult to characterize key elements of the solvation process.
We developed an approach to monitor the time-dependent organization of water clusters, during the final stage of the GPCRs-ligand recognition process, using GPU-accelerated Molecular Dynamics (MD) simulations.

Journal of Chemical Information and Modeling

Perturbation of water's fluid dynamics properties during GPCR-ligand recognition: the human A_{2A} adenosine receptor as a key study.

Davide Sabbadin[#] and Stefano Moro[#]*

[#] *Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università di Padova, via Marzolo 5, Padova, Italy.*

Correspondence to:

Stefano Moro, Ph.D.
Molecular Modeling Section (MMS)
Department of Pharmaceutical and Pharmacological Sciences
University of Padova
Via Marzolo, 5 – 35131 Padova (ITALY)
Tel. +39 049 8275704
FAX +39 049 827 5366
email: stefano.moro@unipd.it

Abbreviations: ARs: Adenosine Receptors; EL2: Second Extracellular Loop; GPCRs: G Protein-Coupled Receptors; GPU: Graphics Processing Unit; WFD maps: Water Fluid Dynamics maps; MD: Molecular Dynamics; n.d.: not determined; NECA: N-Ethyl-5'-Carboxamido Adenosine; POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine; T4E: 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol; T4G: 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine; TM: Transmembrane; ZM 241385: 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino)ethyl)phenol;

Keywords: G Protein-Coupled Receptors; , Membrane Molecular Dynamics, Water molecules, Adenosine Receptors.

ABSTRACT

G protein-coupled receptors (GPCRs) represent the largest family of cell-surface receptors and about one third of the actual targets of clinically used drugs. Recent advances in structural biology described how water molecules play a crucial structural role in GPCRs protein architecture and ligand binding. In the present work, we present an approach to monitor the time-dependent organization of water clusters, during the final stage of the GPCRs-ligand recognition process, using Molecular Dynamics (MD) simulations. We inspect the variation of water's fluid dynamics, mediated by the binding event, with the aim to correlate these results with the binding affinity values of different ligands through the detection of structural water molecules assembly inside the orthosteric binding site of the receptor. The results of this analysis can be shown in a bi-dimensional graph called water's fluid dynamics (WFD) maps. This protocol is valuable when the receptor-ligand complex crystal structure is not yet available, or has not being solved at high resolution, to predict protein "hot-spots" characterized with peculiar shape and electrostatic properties that can play critical role Structure Based Drug Discovery (SBDD).

.

INTRODUCTION

G protein-coupled receptors (GPCRs) are the largest family of cell-surface receptors and represent approximately 3% of the genes in the human genome[1]. They regulate several crucial functions of most cells in the body and receptor dysfunction can lead to a variety of disease conditions[2].

The determination of the rhodopsin crystal structure and, more recently, adrenergic, dopaminergic, histaminergic, opioid and A_{2A} adenosine receptors provides both academia and pharmaceutical companies exceptionally valuable information for a better understanding of the molecular determinants of receptor function and a more reliable rationale for drug design[3]. Very often these aims can be pursued using these structural information in combination with different computational approaches such as molecular docking protocols and molecular dynamic (MD) simulations[4]. In particular, molecular dynamics approaches, adapted to massively parallel computer architectures, have allowed the execution of microsecond-scale standard MD simulations of fully atomistic representations of GPCRs embedded into explicit lipid-water environments[5]. Even if water is the major component of living cells and it has been clearly demonstrated its crucial effect on protein architecture, protein dynamics, ligand binding and protein-mediated ligand transformation (e.g. enzymatic reactions)[6] its role is very often, voluntary or involuntary, omitted.

Focusing our attention on GPCRs, the presence of ordered clusters of water molecules in the proximity of highly conserved motifs in class A GPCRs revealed their structural role in stabilizing intra- and inter-helical interactions[7] and water dynamics revealed to play a pivotal role in both rhodopsin activation and signaling[8,9]. Moreover, the recent high resolution crystal structures of the human A_{2A} adenosine receptors (hA_{2A} ARs) highlighted the active role of water molecules in the ligand-receptor recognition process[10]. In particular, the crucial role of the perturbation of cluster of waters in the ligand-receptor binding process has been recently reinvestigated by Bortolato and collaborators by using different computational methods[11]. In was nicely demonstrated that precise

water modeling is not only an essential requirement for accurate free energy of binding prediction, but also potentially useful in understanding ligand binding kinetics[11].

In the present work, we present an alternative approach to monitor the time-dependent organization of water clusters, during the final stage of the ligand recognition process, using MD simulations. In other words, we would like to inspect the variation of water's fluid dynamics mediated by the binding event with the aim to correlate these results with the binding affinity values of different ligands. To analyze this complex time-dependent process, we elaborated a protocol to detect structural water molecules assembly inside the orthosteric binding site of the receptor. The results of this analysis can be shown in a bi-dimensional graph called water's fluid dynamics (WFD) maps. All membrane MD simulations have been carried out using ACEMD program engineered to run on GPUs[12].

To validate our key study approach, we have selected the human A_{2A} adenosine receptor, that has been recently crystallized with several ligands, both agonists and antagonists, characterized by different receptor binding affinities.

For the present study, we have chosen five crystal structures of the hA_{2A} AR in complex with four strong binders such as 4-(2-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-*a*)(1,3,5)triazin-5-yl-amino)ethyl)phenol, ZM 241385 (pK_D = 9.18 ± 0.00[13], PDB ID: 3EML[14]); 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pK_D = 8.9 ± n.d.[13], PDB ID: 3UZA[15]); 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pK_D = 9.6 ± n.d.[13], PDB ID: 3UZC[15]); NECA: N-Ethyl-5'-Carboxamido Adenosine (pKD = 7.00 ± 0.1[13], PDB ID: 2YDV[16]) and a weaker binder such as caffeine (pK_D = 5.31 ± 0.44[13], PDB ID: 3RFM[13]). A detailed inspection of the caffeine binding mode and hydro-dynamics is repoted in the Supporting Information section. In addition to the exploration of the hydrodynamic profile of most of the crystallographic structures of A_{2A} AR available at on the RCSB Protein Data Bank[17], in order to gather insights on the perturbation of

water's fluid dynamics properties during hA_{2A} AR-ligand recognition, we investigated the time dependent organization of water clusters within the orthosteric binding pocket of the apo-state of the receptor and its ligand-bound state, focusing on various structurally related 1,2,4-triazine derivatives antagonist at the hA_{2A} AR. These findings have great importance since those structures have not been characterized yet by X-RAY spectroscopy. Ligands were obtained by virtual modifications of the 6-(2,6-Dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine and reported in table 1. Most of the hA_{2A} AR antagonists, considered in this study, have been recently synthesized and reported in literature[15].

Table 1

Compound		SlogP(o/w)	Vdw volume (Å)	MW (Da)	pKi	kD
1	1,2,4-triazin-3-amine	-0.34	115.50	96.093	ND	ND
2 (4a)[15]	diphenyl-1,2,4-triazine-3-ammine	3.49	345.45	248.28	6.93	$>1\cdot10^{0}$
3 (4d)[15]	6-(3,5-dimethylphenyl)-5-phenyl-1,2,4-triazin-3-amine	4.20	394.30	276.34	7.67	ND
4	5-phenyl-6-(pyridin-4-yl)-1,2,4-triazin-3-amine	2.26	336.37	249.27	ND	ND
5 (4g)[15] **PDB ID: 3UZA**	6-(2,6-Dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine	2.86	385.23	277.33	8.11	$1.15\cdot10^{-2}$

METHODS

Computational facilities. All computations were performed on a hybrid CPU/GPU cluster. In particular, molecular docking simulations have been carried out using 8 Intel® Xeon® E5620 CPU cluster, whereas membrane molecular dynamics simulation have been performed with a 4 NVIDIA GTX 580 and 2 NVIDIA GTX 680 GPU cluster engineered by Acellera[18]. In the following, the numbering of the amino acids follows the arbitrary scheme by Ballesteros and Weinstein: each amino acid identifier starts with the helix number, followed by the position relative to a reference residue among the most conserved amino acids in that helix, to which the number 50 is arbitrarily assigned[19].

Homology models. The selected five crystal structures (PDB IDs: 3EML, 3UZA, 3UZC, 3RFM and 2YDV) and the FASTA sequence of the hA_{2A} AR (Uniprot ID: P29274) were retrieved from the RCSB PDB database[17] (http://www.rcsb.org) and the UniProtKB/Swiss-Prot[20–22], respectively. The eventual lysozyme portion fused to the receptor as well as co-crystallized ligands and water molecules have been removed before starting the homology modeling procedure. Ionization states and hydrogen positions have been assigned with the "Protonate-3D" tools[23]. Then, to minimize contacts among hydrogens, the structures were subjected to energy minimization with Amber99 force field[24] until the *r.m.s.* of conjugate gradient was < 0.05 $kcal \cdot mol^{-1} \cdot Å^{-1}$, by keeping the heavy atoms fixed at their crystallographic positions. The FASTA sequence was aligned, using Blosum 62 matrix, with the template sequence. Backbone and conserved residues coordinates were copied from the template structure, whereas newly modeled regions and non conserved residues side chains were modeled and energetically optimized by using Amber99 force field until a *r.m.s.* of conjugate gradient < 0.05 $kcal \cdot mol^{-1} \cdot Å^{-1}$ was reached. Missing loop domains were constructed by the loop search method implemented in Molecular Operating Environment (MOE, version 2012.10) program[25] on the basis of the structure of compatible fragments found in the Protein Data Bank[17].

N-terminal and C-terminal were deleted if their lengths exceeded those found in the crystallographic template. The "Protonate-3D" tool[23] was used to appropriately assign ionization states and hydrogen positions to the build models. Protein stereochemistry evaluation was then performed by employing several tools (Ramachandran and χ plots measure j/ψ and χ1/χ2 angles, clash contacts reports) implemented in the MOE suite.

Molecular dynamics. Each ligand-receptor complex was embedded in a 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid bilayer (75x75 Å wide) and placed into the membrane according to the suggested orientation reported in the "Orientations of Proteins in Membranes (OPM)" database[26] for the hA_{2A} AR in complex with the antagonist T4G (PDB ID: 3UZA[15]). Ligand-receptor complexes, where the crystal structure was not available, were obtained by molecular docking using the protocol described previously[4] and physical-chemical descriptors, reported in table 1, were calculated using MOE Suite. Overlapping lipids (within 0.6 Å) were removed upon insertion of the protein. The prepared systems were solvated with TIP3P[27] water using the program Solvate 1.0[28] and neutralized by Na^+/Cl^- counter-ions to a final concentration of 0.154 M. The total number of atoms per system was approximately 35000. Membrane MD simulations were carried out on a GPU cluster with the ACEMD program[12] using the CHARMM27 Force Field[29] and periodic boundaries conditions. Initial parameters for the ligands were derived from the CHARMM General Force Field for organic molecules[30] by using the "paramchem" service[31,32] and were subsequently optimized at the MP2/6-31G* level of theory[33] (consistently with the CHARMM27 Force Field parameterization) by using Gaussian 09[34] and the implemented parameterization tools in the VMD engine[35].

The system was equilibrated using a stepwise procedure. In the first stage, to reduce steric clashes due to the manual setting up of the membrane-receptor system, a 500 steps conjugate-gradient minimization was performed. Then, to allow lipids to reach equilibrium and water molecules to

diffuse into the protein cavity, the system was equilibrated by keeping the positions of protein and ligand atoms restrained for the first 8 ns by a force constant of 1 kcal/mol·Å^2 and then by keeping only the alpha carbon atoms frozen up to 9 ns while gradually reducing the force constant to 0.1 kcal/mol·Å^2. During the equilibration procedure, the temperature was maintained at 298 K using a Langevin thermostat with a low damping constant of 1 ps^{-1}, and the pressure was maintained at 1 atm using a Berendensen barostat. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE[36] algorithm with an integration timestep of 2 fs. Harmonical constraints were then removed during additional 60 ns (NVT ensemble). Long-range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)[37] with grid size rounded to the approximate integer value of cell wall dimensions. A non-bonded cutoff distance of 9 Å with a switching distance of 7.5 Å was used.

Water's fluid dynamics (WFD) maps.

Trajectory analysis, water clustering and water's fluid dynamics (WFD) maps have been generated following the scheme reported in **figure 1** using several functionalities implemented by VMD[35], WORDOM[38], the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC., Gnuplot graphic utility (http://www.gnuplot.info/) and Gromacs tools[39].

For the construction of WFD maps, from the native MD trajectories, the following procedure has been followed:

- The orthosteric binding site has been defined by selecting residues within a range of 5 Å from the bound ligand (including E169[EL2], F168[EL2], His250[6.52], Asn253[6.55], T88[3.36], H278[7.43], E13[1.39] and Y9[1.35]) and a 3D BOX that circumscribe the binding site (**figure 1, panel A**) has been created.

- The built box has been split in a 3D-GRID thus allowing to monitor water diffusion, during the MD simulation, and to localize specific grid cells were water molecules get geometrically trapped hence enabling the characterization of protein hot-spots with peculiar shape and electrostatic properties. Such GRID has been oriented parallel to the z axis (**figure 1, panel A**) in order to make possible further projections of data on the xy plane.

- Each MD trajectory, after system equilibration, has been split in regular time windows of 200 ps in accordance to previous studies on protein hydration[40]. Snapshots of the system coordinates have been saved every 10 ps. Each set of frames, contained in a specific time window, has been processed by calculating the the root mean square fluctuation (r.m.s.f.) of all water molecules contained in the region defined by the originally created box. **Figure 1, panel B** shows that:

    - If the r.m.s.f. of a specific water molecule residue was below 1.4 Å, its position averaged and projected on a 2D-GRID on the xy plane. Averaged coordinates that correspond to the position, within 1.4 Å of a specific water molecule in a specific time window have been recorded into a cumulative PDB file.

    - Otherwise whereas the r.m.s.f. resulted to be above 1.4 Å no projection on a 2D-GRID was made.

- Resulting 2D-GRIDs were overlapped and projected grid cells have been color-coded by normalizing, on a scale from 0% to 100% on a density basis over bulk water, the points inside the projected grid cells thus creating the WFD maps (**figure 1, panel C**).

- The created maps provide an accessible visualization of structural and bulk water distribution inside the human A_{2A} AR orthosteric binding pocket.

Water molecules characterized by a fluctuation below 1.4 Å, within the averaged geometrical position in the selected window of time, define protein hot-spots with peculiar shape and electrostatic properties and show an excellent correlation with the geometrical position and the relative vibrational motion of water molecules experimentally determined in high resolution X-ray structures. This protocol is valuable to predict regions where water molecules can be found in 3D structures where the crystal structure is not yet available or has not being solved at high resolution.

RESULTS AND DISCUSSION

General features of the orthosteric binding site of the hA_{2A} AR.

The binding site of the hA_{2A} AR has been exhaustively described elsewhere[3]. We therefore report here the most relevant receptor-ligand binding features that describe the common interaction pattern for Adenosine Receptor ligands, which are depicted in **figure S1**. This analysis has been carried out by visually inspecting the crystallographic structures of human A_{2A} AR that have been solved and published. The aromatic scaffold of agonists or antagonists is involved an aromatic π-π stacking with the conserved Phe168, located in the second extracellular loop [EL2], and additional hydrophobic contacts with, among others, the Leu249 [6.51] side chain. Strong polar interactions are established with the side chain of the conserved Asn253 [6.55][41], where the role of the hydrogen bond donor in the high-affinity ligands is played, in most cases, by an exocyclic amino group. Moreover in the agonist-bound crystal structure the Thr88 [3.36] side chain forms a hydrogen bond interaction with the nitrogen atom of the acetamide moiety in NECA. This pattern is consistent with the previously reported mutation data, which have been recently reviewed by Cristalli and collaborators[42], showing loss of affinity for the Asn253 [6.55] mutant, as well as with recent mutagenesis data[42] revealing the critical role of Phe168 [EL2] and Leu249 [6.51] for both agonists and antagonists binding and of Thr88 [3.36] for agonist binding.

Recent advances in structural biology allowed to crystallize and resolve a high resolution structure of the A_{2A} AR[10] thus describing how water molecules play a crucial role in bridging protein-ligand interactions by forming a network of hydrogen bonding interaction between Tyr9[1.35] Glu13[1.39] and His278[7.43] and the antagonist ZM241385. Despite this, such information is still missing for other of A_{2A} AR antagonists that posses a better pharmaceutical appeal due to their peculiar ADME properties.

Exploring the water's fluid dynamic profile of hA_{2A} AR in its apo-state. The structure of the hA_{2A} AR in its apo state is still unsolved. In order to depict the possible dynamic organization of water molecules inside the orthosteric binding cleft, the apo-form of the receptor has been embedded in a POPC lipid bilayer and Water's fluid dynamics (WFD) maps have been generated (**figure 2**) as described in detail in the Methods section. The WFD maps highlighted the propensity of specific region of the orthosteric site of the receptor, namely hot-spots, to trap water molecules in a low energetic state that increase their residence time as shown in **figure 2a, 2b** and **2c**. Considering the apo-state of the receptor, this analysis shown that water molecules in proximity of Tyr9 [1.35], Glu13 [1.39], Thr88 [3.36], His250 [6.52] are characterized by an r.m.s.f. < 1.4 Å, over 200 ps, hence suggesting a crucial role in defining the topological and interactive properties of the portion of the orthosteric site. Interestingly, these residues are conserved in all cloned adenosine receptors[43] and it has been demonstrated their involvement in ligand binding process[44–47]. In particular, Glu13 [1.39] and His278 [7.43] have shown to play a critical role in agonist[48] and antagonist recognition and in the allosteric regulation mediated by sodium ions[10]. Curiously, the ε-tautomer (HSE, **figure 2b**) of His278 [7.43] has much less propensity to coordinate water molecules compared to its δ-tautomer (HSD, **figure 2a**) and protonated state (HSP, **figure 2c**).

Besides Tyr9 [1.35] Glu13 [1.39] and His278 [7.43] side chains surrounding volume (3 Å), the remaining regions of the orthosteric binding site are statistically occupied by a water molecules with a residence time compare to the water in the bulk liquid. These low residence time water molecules are more likely prone to replacement by ligand binding at no energetic cost.

Hot-spots reveals that the trapped water molecules re-shape of the orthosteric binding pocket accessible volume by the ligand at no enthalpic cost.

Ligand binding to the receptor is energetically driven by the Gibbs free energy equation. Upon binding the loss of solvent interactions within ligand and protein alone lead to an unfavorable

enthalpic contribution, whereas these structured water molecules are released to bulk solvent upon ligand binding, which leads to a favorable increase in entropy[49–51] and the displacement of unfavorable waters by the ligand, replacing them with groups complementary to the protein surface, represents a crucial driving force for protein−ligand binding[11].

The release of a highly ordered water molecule from the active site to bulk solution theoretically results in an entropic gain of 7 cal mol^{-1} K^{-1}[52].

Theoretically, if the ligand upon binding removes just bulk waters, ligand-receptor recognition can be done at no enthalpic cost. Calculating the volume of the binding pocket accessible by bulk solvent molecules we found it to be ~30% smaller than calculating the accessible volume using only the protein structure alone (data not shown). This concept can be taken into account when designing ligands that needs to have specific physical-chemical properties, especially if targeting the Central Nervous System (CNS), since the physical properties in general have a smaller range than the criteria defined by the Lipinski rule of five[53].

The generated water network maps and 3D structure file (see supporting material) can be used in tandem with other approaches in order to enable intelligent scaffold replacement, or other chemical modifications, that do not perturb crucial water molecules thus lowering toxicity and maintaining potency and selectivity profile. This approach could simplify the discovery of a new non-furan, non-xanthine and relatively polar hA_{2A} AR targeting agent characterized with an eased path to approval.

Exploring ligand bound-hA_{2A} AR hydrodynamic profile.

As stated previously, several crystallographic structures of the human adenosine A_{2A} receptor, in complex with different agonists and antagonists, have been solved and released. The ARs physiological agonist adenosine (PDB ID: 2YDO[16]), its N-ethyl-5′carboxamide derivative, NECA,

(PDB ID: 2YDV[16]) and the high affinity agonist UK-432097, 6-(2,2-diphenylethylamino)-9-[(2R,3R,4S,5S)-5-(ethylcarbamoyl)-3,4-dihydroxy-oxolan-2-yl]-N-[2-[(1-pyridin-2-ylpiperi-din-4-yl)carbamoylamino]ethyl]purine-2-carboxamide, (PDB ID: 3QAK[54]) have been co-crystallized with the human A_{2A} AR. Moreover, antagonists belonging to different chemical families, have been also co-crystallized with the human A_{2A} AR. In particular; the high affinity antagonist (4-(2-[7-amino-2-(2-furyl) [1,2,4]-triazolo[2,3-α] [1,3,5]triazin-5-ylamino]ethyl)phenol, better known as ZM241385, is in complex with the human Adenosine A2A receptor/T4 lysozyme chimera (PDB ID: 3EML[14]) and in complex with other hA_{2A} AR mutants/chimeras (PDB ID: 3PWH[13] PDB ID: 3VGA[55], PDB ID: 3VG9[55] and PDB ID: 4EIY[10]). Finally, xantine derivatives such as the N-(2-aminoethyl)-2-[4-(2,6-dioxo-1,3-dipropyl-2,3,6,7-tetrahydro-1H-purin-8-yl)phenoxy]acetamide (PDB ID: 3REY[13]) and the very well known caffeine (PDB ID: 3RFM[13]) have been co-crystallized with A_{2A} adenosine receptor. Recently, ARs structural information have been furthermore enriched by the co-crystallization of 1,2,4-triazine derivatives such as the 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine (PDB ID: 3UZA[15]) and the 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (PDB ID: 3UZC[15]) with a thermostabilised human adenosine A_{2A} receptor. Expected changes in the hydrodynamic profile of the orthosteric binding pocket between ligand-bound and apo-state of hA_{2A} AR revealed the role of water molecules in mediating protein-ligand binding.

The heat-maps corresponding to the **ZM241385-hA_{2A} AR complex** hydrodynamic profile (**figure 3a**) highlight the presence of an enriched arrangement of water molecules that bridges Tyr9 [1.35] Glu13 [1.39] and His278 [7.43] to the triazolotriazine core of the antagonist compared to the solvation profile of the apo state of the receptor. Additionally the hot-spot located proximity of Asn253 [6.55] and Glu169 [EL2] highlight that the exocyclic nitrogen of the antagonist interactions with the cited side chains are stabilized by a stable cluster of water molecules. ZM241385 binding also induces a re-arrangement of water molecules that are bound to Thr88 [3.36] side chain (**figure 3e**).

There is an excellent correlation with the experimentally determined high resolution X-ray structures as show in **figure 3b.** In fact, the arrangement of crystallized water molecules in the crystal structure of the chimeric protein of A_{2A} AR-BRIL in complex with ZM241385 at 1.8A resolution (PDB ID: 4EIY[10]) and their associated B-factor values reflect the WFD map obtained in our work. The cluster of water molecules (W2527, W2521, W2520, W2584, W2585, W2524, W2525) that interact with the triazolotriazine core of the antagonist and Tyr9 [1.35] Glu13 [1.39] and His278 [7.43] are located in geometrical positions were enriched cluster of structural water molecules have been predicted to be present in the same region they have been found experimentally (**figure 3a**). Water W2517 interact with Asn253 [6.55], Glu169 [EL2] and the exocyclic nitrogen of ZM241385. Water W2572, W2668, W2583 and W2541 surround the phenol moiety of the antagonist and Glu169 [EL2]. Temperature factor values of water molecules that surround the antagonist in the crystallographic structure range from 16.75 to 47.59. These molecules representations in panel **b c** and **d** of **figure 3** were colored following the same color code used to generate the hydrodynamic heat maps thus finding an extremely similar pattern between the relative vibrational motion of the selected parts of the structure and the reduced ability to fluctuate over time calculated on MD trajectories. Lower resolution crystal structures of A_{2A} AR in complex with the antagonist ZM241385 (PDB ID: 3EML[14] – **figure 3c**; PDB ID: 3VG9[55] – **figure 3d**) reveal a similar solvation pattern but in a lower level of detail.

In the agonist bound (**NECA**) **hA_{2A} AR complex** the WFD map (depicted in **figure 4a**) is highly correlated with the geometrical position and B-factor values of water molecules co-crystallized in the XRAY crystal structure of the thermostabilized human A_{2A} AR with NECA bound (PDB ID: 2YDV[16]). Water molecules W2017 and W2027 bridge the exocyclic nitrogen of the adenine ring to the Glu169 [EL2] and Asn253 [6.55] side chains and the arrangement of water molecules that surrounds Thr88 [3.36] and His278 [7.43] in the apo-state of hA_{2A} AR are displaced by the direct interaction of the acetamide moiety of NECA with the hydroxyl group of side chain. As reported for

the antagonist ZM241385 Tyr9 [1.35] and Glu13 [1.39] represent an hot spot that bridges TM 1 to the aromatic scaffold of NECA through water molecules W2026, W2002 and W2001.

Despite the binding mode of some of the potent **1,2,4-triazine derivatives**, antagonist at the human A_{2A} AR, have been revealed by XRAY crystallography[15] , information about the role of water molecules in ligand binding is still lacking. WFD maps obtained in this study (**figure 5a, 5b**) highlighted that the strong interaction between Asn253 [6.55] and Glu169 [EL2] side chains and the exocyclic nitrogen bound the 1,2,4-triazine core is stabilized by ordered water molecules. The Nitrogen atom of the pyridyl moiety in 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4- triazin-3-amine (T4G) interacts with His278 [7.43] by a ordered cluster of water molecules (**figure 5a**) and, as previously seen, induces a re-arrangement of water molecules that are bound to Thr88 [3.36] side chain (**figure 5e, 5d**). Upon binding of the Chloro-phenol derivative (T4E) the ordered cluster of water molecules bound to His278 [7.43] and Glu13 [1.39] are released into the bulk solvent, leading to an increase in entropy, favourable for ligand binding. Interestingly an ordered arrangement of water molecules, on both 1,2,4-triazine derivatives, around the nitrogen 1 and 2 of the aromatic core, favourably contributes to the free energy of binding of this class of hA_{2A} AR antagonists.

Water Fluid Dynamic Maps as supporting tool for an effective in silico drug discovery strategy.

The 1,2,4-triazine-3-amine (**compound 1, table 1**) is predicted to be the most polar of the considered set of molecules (**table 1**) and is characterized by a small van der Waals volume (115.50 $Å^3$). There is no proven binding of this molecules to the human A_{2A} AR and the MD simulation, of a possible ligand-receptor complex, highlighted the unbinding (r.m.s.d.>20Å) of the originally docked compound from the orthosteric binding pocket. The analysis of the hydrodynamic profile of the orthosteric binding pocket (**figure 6b**) highlight that, upon ligand-receptor interaction, a loss of the water molecules (**wA** cluster, **figure 6a**), that are trapped in a low energetical state and

bound to specific hot-spots of the receptor such as the triad Tyr9 [1.35] Glu13 [1.39] and His278 [7.43], occur thus showing the propensity of bulk water to solvate the polar fragment-like molecule and facilitating it's unbinding from hA_{2A} AR.

The substitution of the two hydrogens in C5 and C6 of the triazine core with two phenyl groups increases dramatically the volume of the triazine derivative, the commercially available diphenyl-1,2,4-triazine-3-ammine (**compound 2, table 1**) to 345.45 $Å^3$. The phenyl substituent from the C5 position of the triazine ring occupies the hydrophobic pocket enclosed by Leu84 [3.32], Leu85 [3.33], Met177 [5.38], Trp246 [6.48] ,Leu249 [6.51], and His250 [6.52] (data not shown). The second phenyl substituent from the C6 carbon of the 1,2,4-triazine-3-ammine scaffold pointing toward a hydrophobic region defined by Ala63 [2.61] and Ile66 [2.64] and the His278 [7.43]. The chemical modification that lead to the dimethyl-phenyl derivative (**compound 3, table 1**) provides enhanced surface complementarity between ligand and receptor, thus improving ligand binding. The hydrodynamic maps show that, upon ligand binding, a cluster of water molecules (**wB** cluster, **figure 6c and d**) mediates interactions between Glu169 [EL2], Asn253 [6.55] and the exo-cyclic Nitrogen bound to the triazine ring.

Interestingly the hydrophobic moieties of **compounds 2** and **3** that point towards Tyr9 [1.35] Glu13 [1.39] and His278 [7.43] unhinge the water molecules network that is present in the neighboring region prior ligand binding.

The insertion of a pyridyl substituent from the C6 position of the triazine ring (**compound 4, table 1**) results in the creation of two "non-bulk" structural water molecules arrangements (**wA** and **wB** clusters, figure **6e**), that mediate interactions between Glu169 [EL2], Asn253 [6.55] and the exo-cyclic Nitrogen bound to the triazine ring and bridge polar interactions between the pyridyl Nitrogen and Glu13 [1.39] and His278 [7.43].

The combination of such modifications that lead from **compound 2** to the **compound 5** (**table 1**), such the 3-5-methylation and the insertion of a nitrogen atom in the para position of the substituent from the C6 position of the triazine ring, increased affinity to the hA_{2A} AR of about an order of magnitude and a peculiar arrangement of water molecules, around the ligand, that are characterized by low geometrical fluctuation during MD simulations are depicted in **figure 6f** and in figure **5a**.

The three emerging stable clusters of water molecules namely **wA, wB** and **wC** (**figure 6f**) are found to interact with both receptor and ligand thus contributing to its low fluctuation inside the orthostetic binding site and perhaps providing a rational basis that can explain the slower off-rate receptor kinetics (~ two orders of magnitude)[15] of **compound 5** (1.15×10^{-2}) than the other ones reported in **table 1**.

CONCLUSIONS

All of the data displayed on the maps is based on geometrical information collected from fully atomistic MD simulations of ligand-receptor complexes, or apo-state of the receptor, embedded in an explicit lipid-water environment, thus taking advantage of the thermodynamic accuracy, at the molecular level, of MD simulations. This approach is versatile and facilitate GPU-driven research by letting scientists decide which bio-molecular simulations package, that could be engineered to run on GPUs, suit their needs. Eventually we also believe that the proposed strategy can be extended to other GPCRs as well as to homology models.

ACKNOWLEDGMENT

The molecular modeling work coordinated by S.M. has been carried out with financial support of the University of Padova, Italy, and the Italian Ministry for University and Research (MIUR),

Rome, Italy. S.M. is also very grateful to Chemical Computing Group and Acellera Ltd. for the scientific and technical partnership.

ADDITIONAL CONTENT

Supporting Information.

A discussion on caffeine hydration profile is reported on the Supplementary materials.

This material is available free of charge via the Internet at http://pubs.acs.org

REFERENCES

(1) Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 639–650.

(2) Sodhi, A.; Montaner, S.; Gutkind, J. S. Viral hijacking of G-protein-coupled-receptor signalling networks. *Nature Reviews Molecular Cell Biology* **2004**, *5*, 998–1012.

(3) Jacobson, K. A.; Costanzi, S. New insights for drug design from the X-ray crystallographic structures of G-protein-coupled receptors. *Mol. Pharmacol.* **2012**, *82*, 361–371.

(4) Inamdar, G. S.; Pandya, A. N.; Thakar, H. M.; Sudarsanam, V.; Kachler, S.; Sabbadin, D.; Moro, S.; Klotz, K.-N.; Vasu, K. K. New insight into adenosine receptors selectivity derived from a novel series of [5-substituted-4-phenyl-1,3-thiazol-2-yl] benzamides and furamides. *Eur J Med Chem* **2013**, *63*, 924–934.

(5) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13118–13123.

(6) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108*, 74–108.

(7) Pardo, L.; Deupi, X.; Dölker, N.; López-Rodríguez, M. L.; Campillo, M. The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *Chembiochem* **2007**, *8*, 19–24.

(8) Angel, T. E.; Gupta, S.; Jastrzebska, B.; Palczewski, K.; Chance, M. R. Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14367–14372.

(9) Jastrzebska, B.; Palczewski, K.; Golczak, M. Role of bulk water in hydrolysis of the rhodopsin chromophore. *J. Biol. Chem.* **2011**, *286*, 18930–18937.

(10) Liu, W.; Chun, E.; Thompson, A. A.; Chubukov, P.; Xu, F.; Katritch, V.; Han, G. W.; Roth, C. B.; Heitman, L. H.; IJzerman, A. P.; Cherezov, V.; Stevens, R. C. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* **2012**, *337*, 232–236.

(11) Bortolato, A.; Tehan, B. G.; Bodnarchuk, M. S.; Essex, J. W.; Mason, J. S. Water Network Perturbation in Ligand Binding: Adenosine A2A Antagonists as a Case Study. *J Chem Inf Model* **2013**, *53*, 1700–1713.

(12) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.

(13) Doré, A. S.; Robertson, N.; Errey, J. C.; Ng, I.; Hollenstein, K.; Tehan, B.; Hurrell, E.; Bennett, K.; Congreve, M.; Magnani, F.; Tate, C. G.; Weir, M.; Marshall, F. H. Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* **2011**, *19*, 1283–1293.

(14) Jaakola, V.-P.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; Stevens, R. C. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.

(15) Congreve, M.; Andrews, S. P.; Doré, A. S.; Hollenstein, K.; Hurrell, E.; Langmead, C. J.; Mason, J. S.; Ng, I. W.; Tehan, B.; Zhukov, A.; Weir, M.; Marshall, F. H. Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J. Med. Chem.* **2012**, *55*, 1898–1903.

(16) Lebon, G.; Warne, T.; Edwards, P. C.; Bennett, K.; Langmead, C. J.; Leslie, A. G. W.; Tate, C. G. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* **2011**, *474*, 521–525.

(17) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.* **2000**, *28*, 235–242.

(18) Acellera http://www.acellera.com/.

(19) Ballesteros, J. A.; Weinstein, H. [19] Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. In *Methods in Neurosciences*; Stuart C. Sealfon, Ed.; Academic Press, 1995; Vol. Volume 25, pp. 366–428.

(20) The Universal Protein Resource (UniProt) in 2010. *Nucleic Acids Res.* **2010**, *38*, D142–148.

(21) Jain, E.; Bairoch, A.; Duvaud, S.; Phan, I.; Redaschi, N.; Suzek, B. E.; Martin, M. J.; McGarvey, P.; Gasteiger, E. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics* **2009**, *10*, 136.

(22) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

(23) Labute, P. Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **2009**, *75*, 187–205.

(24) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry* **2000**, *21*, 1049–1074.

(25) Molecular Operating Environment http://www.chemcomp.com/.

(26) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: orientations of proteins in membranes database. *Bioinformatics* **2006**, *22*, 623–625.

(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79*, 926–935.

(28) Helmut Grubmüller and Volker Groll Solvate http://www.mpibpc.mpg.de/home/grubmueller/downloads/solvate/index.html.

(29) MacKerell, A. D., Jr; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–265.

(30) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010**, *31*, 671–690.

(31) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**.

(32) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**.

(33) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chemical Physics Letters* **1988**, *153*, 503–506.

(34) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, revision B.01; Gaussian Inc.: Wallingford, CT, 2009.

(35) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14*, 33–38, 27–28.

(36) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *Journal of Computational Chemistry* **2001**, *22*, 501–508.

(37) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(38) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* **2011**, *32*, 1183–1194.

(39) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005**, *26*, 1701–1718.

(40) Otting, G.; Liepinsh, E.; Wüthrich, K. Protein hydration in aqueous solution. *Science* **1991**, *254*, 974–980.

(41) Jaakola, V.-P.; Lane, J. R.; Lin, J. Y.; Katritch, V.; Ijzerman, A. P.; Stevens, R. C. Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: identification and characterization of essential amino acid residues. *J. Biol. Chem.* **2010**, *285*, 13032–13044.

(42) Cristalli, G.; Lambertucci, C.; Marucci, G.; Volpini, R.; Dal Ben, D. A2A adenosine receptor and its modulators: overview on a druggable GPCR and on structure-activity relationship analysis and binding requirements of agonists and antagonists. *Curr. Pharm. Des.* **2008**, *14*, 1525–1552.

(43) Floris, M.; Sabbadin, D.; Medda, R.; Bulfone, A.; Moro, S. Adenosiland: walking through adenosine receptors landscape. *Eur J Med Chem* **2012**, *58*, 248–257.

(44) Jiang, Q.; Van Rhee, A. M.; Kim, J.; Yehle, S.; Wess, J.; Jacobson, K. A. Hydrophilic side chains in the third and seventh transmembrane helical domains of human A2A adenosine receptors are required for ligand recognition. *Mol. Pharmacol.* **1996**, *50*, 512–521.

(45) Jiang, Q.; Lee, B. X.; Glashofer, M.; Van Rhee, A. M.; Jacobson, K. A. Mutagenesis reveals structure-activity parallels between human A2A adenosine receptors and biogenic amine G protein-coupled receptors. *J. Med. Chem.* **1997**, *40*, 2588–2595.

(46) Kim, J.; Wess, J.; Van Rhee, A. M.; Schöneberg, T.; Jacobson, K. A. Site-directed mutagenesis identifies residues involved in ligand recognition in the human A2a adenosine receptor. *J. Biol. Chem.* **1995**, *270*, 13987–13997.

(47) Kim, S.-K.; Gao, Z.-G.; Van Rompaey, P.; Gross, A. S.; Chen, A.; Van Calenbergh, S.; Jacobson, K. A. Modeling the adenosine receptors: comparison of the binding domains of A2A agonists and antagonists. *J. Med. Chem.* **2003**, *46*, 4847–4859.

(48) IJzerman, A. P.; Von Frijtag Drabbe Künzel, J. K.; Kim, J.; Jiang, Q.; Jacobson, K. A. Site-directed mutagenesis of the human adenosine A2A receptor. Critical involvement of Glu13 in agonist recognition. *Eur. J. Pharmacol.* **1996**, *310*, 269–272.

(49) Dunitz, J. D. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.* **1995**, *2*, 709–712.

(50) Poornima, C. S.; Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput. Aided Mol. Des.* **1995**, *9*, 500–512.

(51) Helms, V. Protein dynamics tightly connected to the dynamics of surrounding and internal water molecules. *Chemphyschem* **2007**, *8*, 23–33.

(52) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670.

(53) Pajouhesh, H.; Lenz, G. R. Medicinal Chemical Properties of Successful Central Nervous System Drugs. *NeuroRx* **2005**, *2*, 541–553.

(54) Xu, F.; Wu, H.; Katritch, V.; Han, G. W.; Jacobson, K. A.; Gao, Z.-G.; Cherezov, V.; Stevens, R. C. Structure of an agonist-bound human A2A adenosine receptor. *Science* **2011**, *332*, 322–327.

(55) Hino, T.; Arakawa, T.; Iwanari, H.; Yurugi-Kobayashi, T.; Ikeda-Suno, C.; Nakada-Nakura, Y.; Kusano-Arai, O.; Weyand, S.; Shimamura, T.; Nomura, N.; Cameron, A. D.; Kobayashi, T.; Hamakubo, T.; Iwata, S.; Murata, T. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature* **2012**, *482*, 237–240.

FIGURE LEGENDS

Figure 1. Workflow of the Water's fluid dynamics (WFD) maps. Construction process. Panel **A**: 3D-BOX definition that circumscribe the orthosteric binding site; panel **B**: 2D-GRID projection of water molecules that get geometrically trapped during MD simulations; panel **C**: 2D-GRIDs were overlap and WFD simplified representation.

Figure 2.

Probing His278 [7.43] tautomers effect on the hydro-dynamic profile of the apo-state of hA2A AR. Panel **A**: δ-tautomer (HSD); panel **B**: ε-tautomer (HSE); panel **C**: fully protonated state (HSP). Region colored in white-light green define bulk water occupancy. Yellow-blue areas define protein "hot-spots" where transient water molecules get trapped during MD simulations. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed.

Figure 3.

Water's Fluid Dynamic map for ZM241385-hA_{2A} AR complex (panel **A**); panels **B-C-D**: vibrational motion of water molecules experimentally determined in high resolution X-ray structures; panel **E**: Ordered water molecules enrichment in comparison to the apo-state of hA_{2A} AR. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed.

Figure 4.

Water's Fluid Dynamic map for NECA-hA_{2A} AR complex (panel **A**); panel **B**: vibrational motion of water molecules experimentally determined in high resolution X-ray structures; panel **C**: Ordered water molecules enrichment in comparison to the apo-state of hA_{2A} AR. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed.

Figure 5.

Water's Fluid Dynamic map for T4G-hA_{2A} AR complex (panel **A**); panels **B:** Water's Fluid Dynamic map for T4E-hA_{2A} AR; panel **C-D**: Ordered water molecules enrichment in comparison between the apo-state of hA_{2A} AR and WFD of hA_{2A} AR bound-T4G and T4E, respectively. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed.

Figure 6.

Combined Water's Fluid Dynamic maps and r.m.s.d. profiles for compound 1 to compound 5 (panel **B** to **F**) into the hA_{2A} AR binding pocket. *1 ns time window. Hydro-dynamic profile of the apo-state of hA2A A is depicted in panel **A**. Receptors are viewed from the membrane side facing TM6 and TM7. Side chains of the amino acids crucial for ligand binding are displayed as gray sticks. Hydrogen atoms are not displayed.

Table of Contents Graphic.

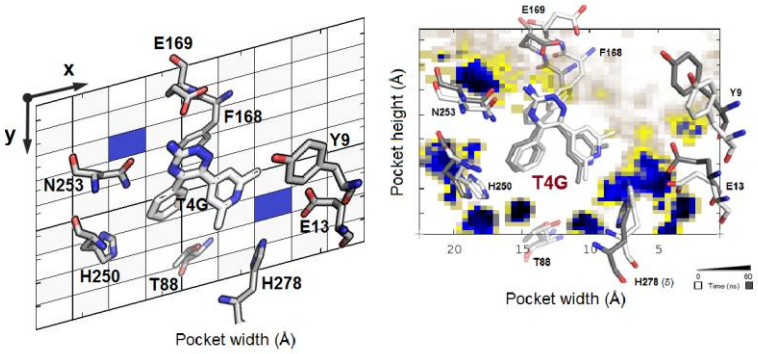


Figure 1.

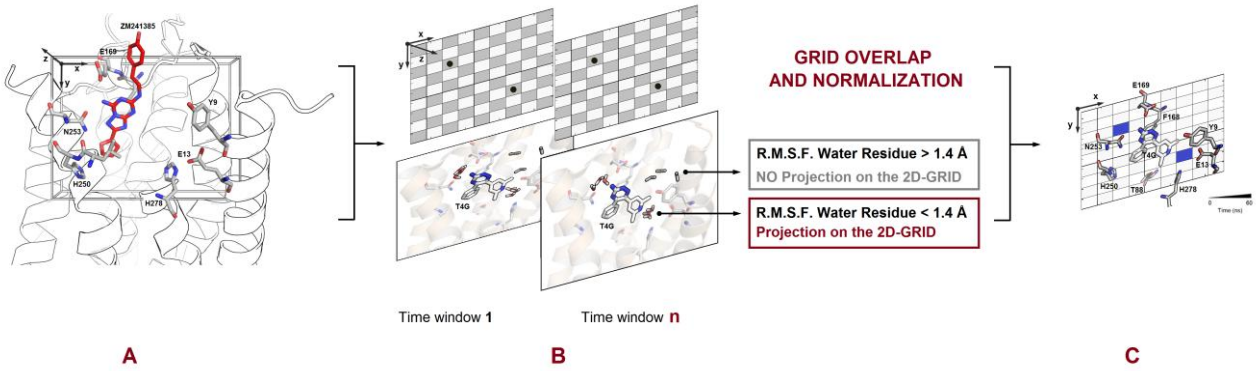

ZM241385
E169
Y9
N253
E13
H250
H278
T4G
GRID OVERLAP
AND NORMALIZATION
R.M.S.F. Water Residue > 1.4 Å
NO Projection on the 2D-GRID
R.M.S.F. Water Residue < 1.4 Å
Projection on the 2D-GRID
F168
T88
Time window 1
Time window n
A
B
C

Figure 2.

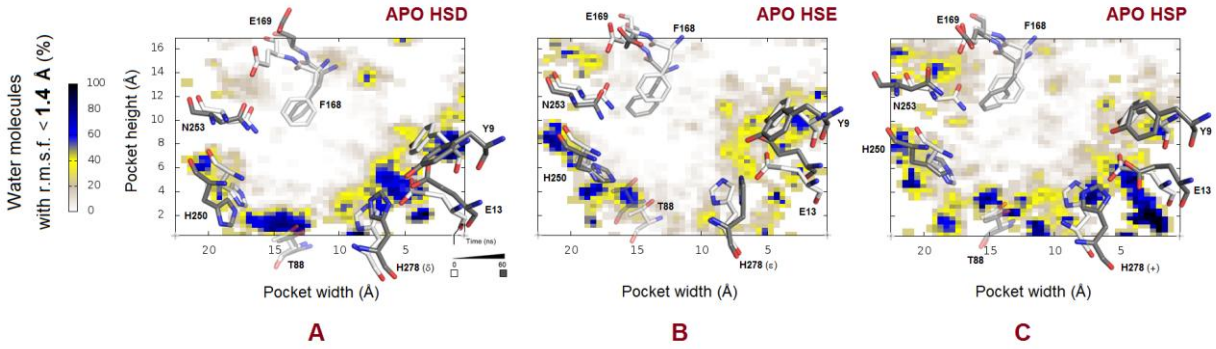

Water molecules with r.m.s.f. < 1.4 Å (%)
100
80
60
40
20
0
Pocket height (Å)
APO HSD
APO HSE
APO HSP
E169
F168
N253
Y9
H250
E13
T88
H278 (δ)
H278 (ε)
H278 (+)
Pocket width (Å)
A
B
C

Figure 3.

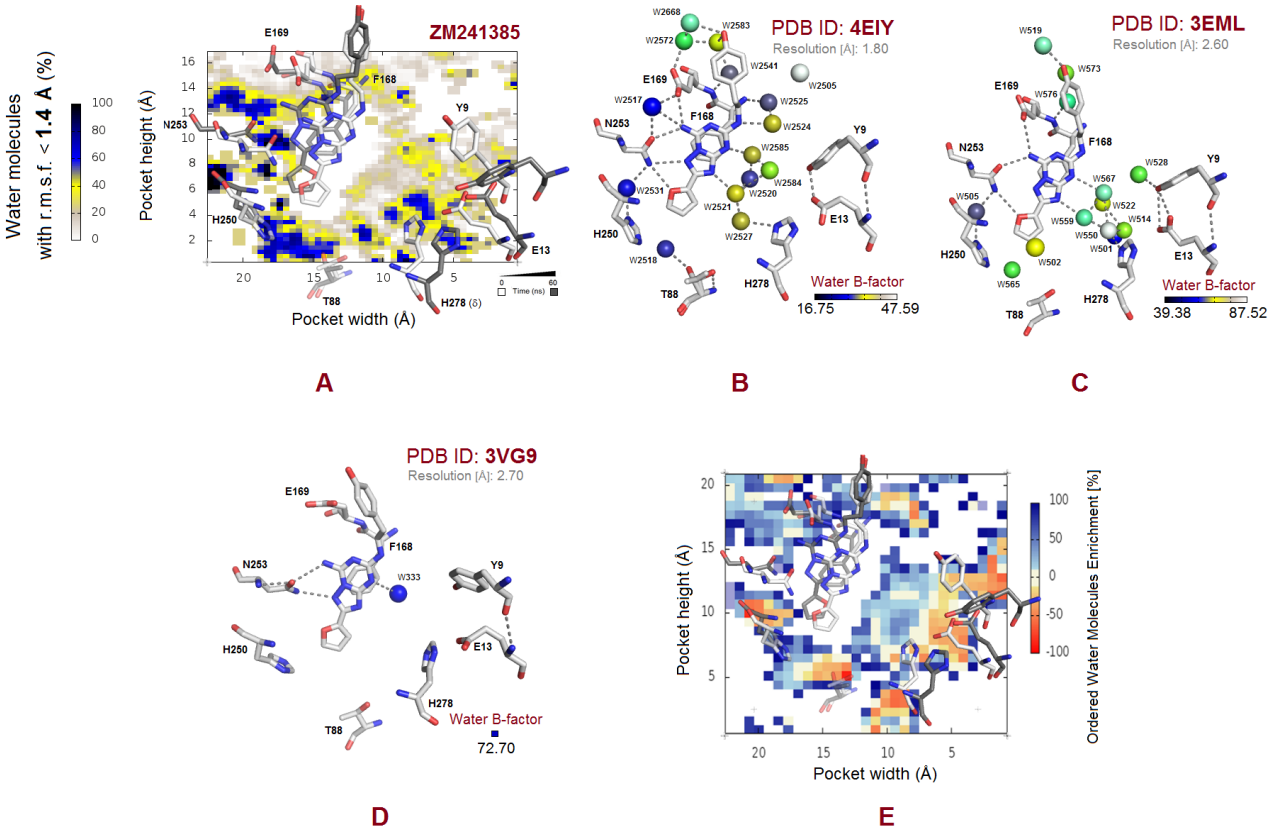

Water molecules with r.m.s.f. < 1.4 Å (%)
100
80
60
40
20
0
ZM241385
Pocket height (Å)
Pocket width (Å)
E169
F168
Y9
N253
H250
E13
T88
H278 (δ)
Time (ns)
A
PDB ID: 4EIY
Resolution [Å]: 1.80
Water B-factor
16.75
47.59
B
PDB ID: 3EML
Resolution [Å]: 2.60
Water B-factor
39.38
87.52
C
PDB ID: 3VG9
Resolution [Å]: 2.70
W333
Water B-factor
72.70
D
Ordered Water Molecules Enrichment [%]
E

Figure 4.

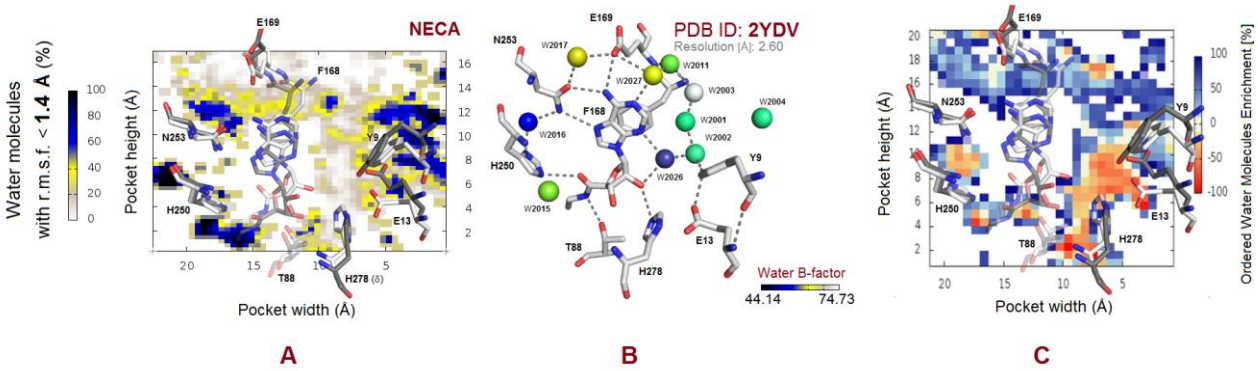

Water molecules with r.m.s.f. < 1.4 Å (%)
100
80
60
40
20
0
NECA
E169
F168
N253
Y9
H250
E13
T88
H278 (δ)
Pocket height (Å)
Pocket width (Å)
PDB ID: 2YDV
Resolution [Å]: 2.60
W2017
W2011
W2027
W2003
W2004
W2016
W2001
W2002
W2026
W2015
Water B-factor
44.14
74.73
Ordered Water Molecules Enrichment [%]
A
B
C

Figure 5.

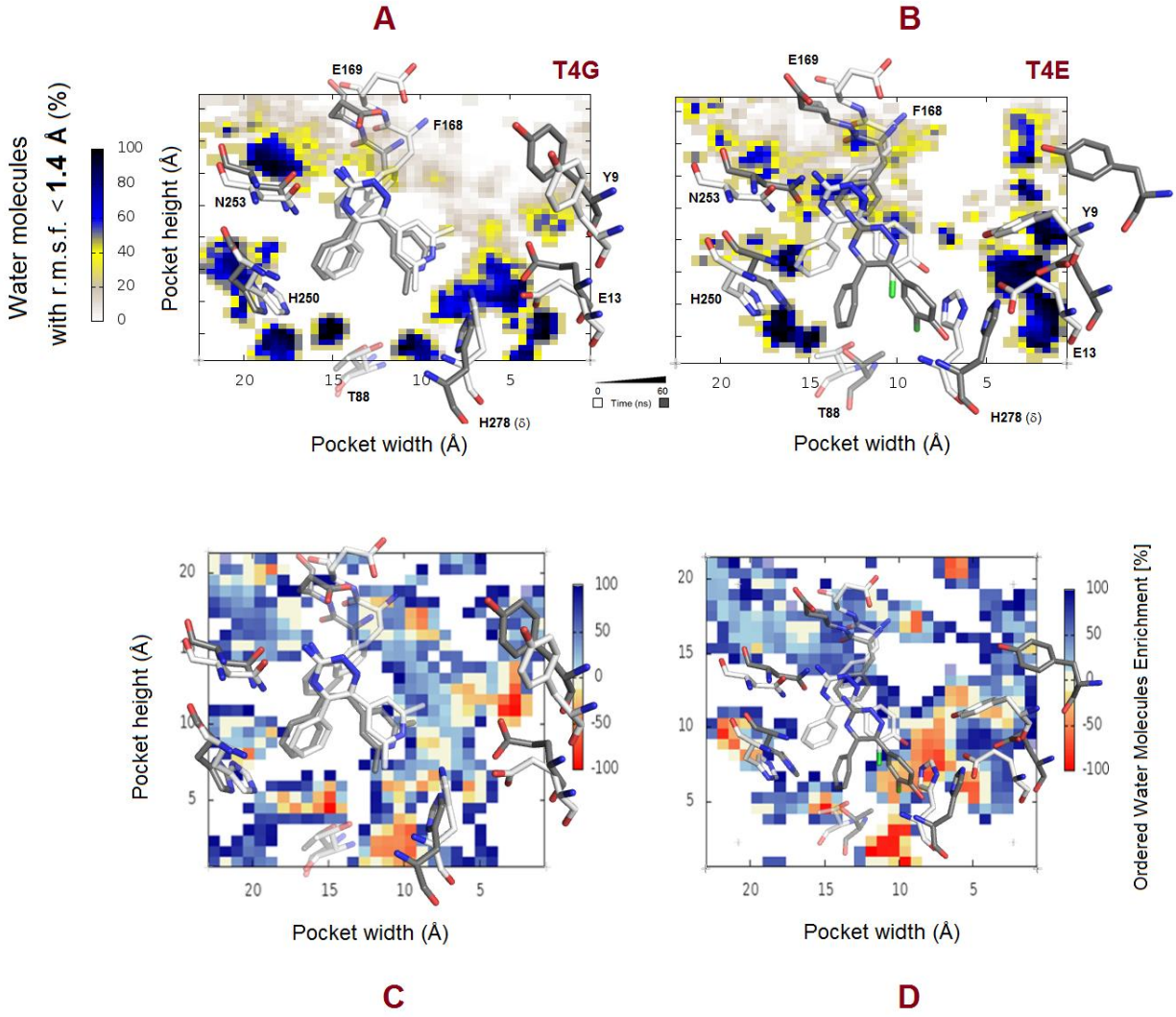
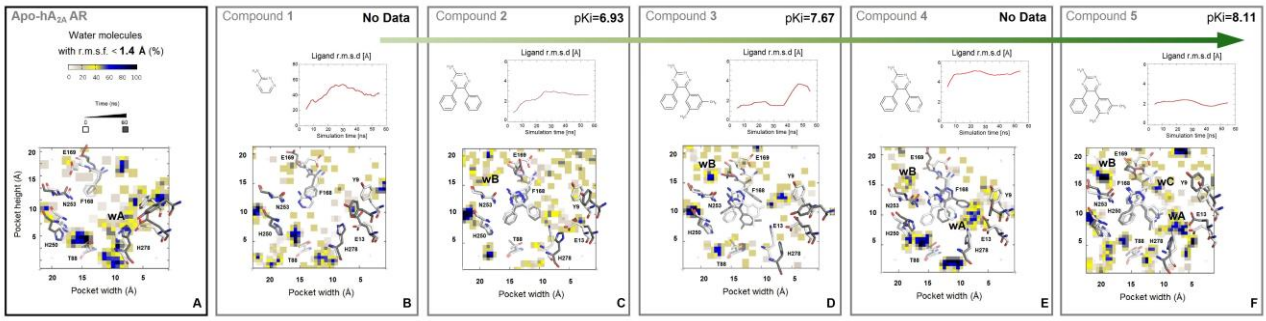

A
B
T4G
T4E
E169
F168
N253
Y9
H250
E13
T88
H278 (δ)
Water molecules with r.m.s.f. < 1.4 Å (%)
100
80
60
40
20
0
Pocket height (Å)
Pocket width (Å)
20
15
10
5
0
60
Time (ns)
C
D
100
50
0
-50
-100
Ordered Water Molecules Enrichment [%]

Figure 6.


Apo-hA2A AR
Water molecules
with r.m.s.f. < 1.4 Å (%)
Pocket height (Å)
Pocket width (Å)
wA
A
Compound 1
No Data
Ligand r.m.s.d [Å]
B
Compound 2
pKi=6.93
wB
C
Compound 3
pKi=7.67
D
Compound 4
No Data
E
Compound 5
pKi=8.11
wC
F

Davide Sabbadin's publications

1. **Sabbadin, D.**, Moro, *Supervised Molecular Dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale* (2014). Journal of Chemical Information and Modeling. In Press.

2. **Sabbadin, D.**, Moro, S. Hydrodynamic 2D/3D-mapping of protein solvation profile using GPU-driven all-atoms Molecular Dynamics (2013). Article in preparation.

3. **Sabbadin, D.**, Walker, R. C., Taylor, S. S. *Accelerated Molecular Dynamics reveal the dynamics of insertion of PKA C-subunit bound Myristoyl group into the hydrocarbon region of a lipid bilayer* (2013) Protein Science. Article in preparation.

4. Cagide, F., Gaspar, A., Reis, J., Uriarte, E., **Sabbadin, D.**, Moro S., Kachler, Klotz, K.-N., Borges F., *Development of A_3 Adenosine receptor ligands based on a chromone-2-carboxamide scaffold: insights into the structure-affinity relationships* (2013) Journal of Medicinal Chemistry. Article in preparation.

5. **Sabbadin, D.**, Ciancetta, A., Moro, S. *Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A_{2A} adenosine receptor as a key study* (2013) Journal of Chemical Information and Modeling.

6. Fanton, M., Floris, M., Cristiani, A., Olla, S., Medda, R., **Sabbadin, D.**, Bulfone, A., Moro, S. MMsDusty: *An alternative InChI-based tool to minimize chemical redundancy* (2013) Molecular Informatics, 32 (8), pp. 681-684.

7. Raffaello, A., De Stefani, D., **Sabbadin, D.**, Teardo, E., Merli, G., Picard, A., Checchetto, V., Moro, S., Szabò, I., Rizzuto, R. *The mitochondrial calcium uniporter is a multimer that can include a dominant-negative pore-forming subunit* (2013) EMBO Journal. In Press.

8. Inamdar, G.S., Pandya, A.N., Thakar, H.M., Sudarsanam, V., Kachler, S., **Sabbadin, D.**, Moro, S., Klotz, K.-N., Vasu, K.K. *New insight into adenosine receptors selectivity derived from a novel series of [5-substituted-4-phenyl-1,3-thiazol-2-yl] benzamides and furamides* (2013) Eu-

ropean Journal of Medicinal Chemistry, 63, pp. 924-934.

9. **Sabbadin, D.**, Floris, M., Medda, R., Bulfone, A., Moro, S. *Adenosiland: Walking through adenosine receptors landscape* (2012) European Journal of Medicinal Chemistry, 58, pp. 248-257.

10. Federico, S., Ciancetta, A., **Sabbadin, D.**, Paoletta, S., Pastorin, G., Cacciari, B., Klotz, K.N., Moro, S., Spalluto, G. *Exploring the directionality of 5-substitutions in a new series of 5-alkylaminopyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine as a strategy to design novel human A3 adenosine receptor antagonists* (2012) Journal of Medicinal Chemistry, 55 (22), pp. 9654-9668.

11. Colotta, V., Lenzi, O., Catarzi, D., Varano, F., Squarcialupi, L., Costagli, C., Galli, A., Ghelardini, C., Pugliese, A.M., Maraula, G., Coppi, E., Pellegrini-Giampietro, D.E., Pedata, F., **Sabbadin, D.**, Moro, S. *3-Hydroxy-1H-quinazoline-2,4-dione derivatives as new antagonists at ionotropic glutamate receptors: Molecular modeling and pharmacological studies* (2012) European Journal of Medicinal Chemistry, 54, pp. 470-482.

12. Coluccia, A., **Sabbadin, D.**, Brancale, A. *Molecular modelling studies on Arylthioindoles as potent inhibitors of tubulin polymerization* (2011) European Journal of Medicinal Chemistry, 46 (8), pp. 3519-3525.

Bibliography

[1] Kristen L Pierce, Richard T Premont, and Robert J Lefkowitz. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.*, 3(9), September 2002.

[2] Kurt Kristiansen. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of g-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.*, 103(1), July 2004.

[3] Akrit Sodhi, Silvia Montaner, and J. Silvio Gutkind. Viral hijacking of g-protein-coupled-receptor signalling networks. *Nature Reviews Molecular Cell Biology*, 5(12):998–1012, December 2004.

[4] J Drews. Drug discovery: a historical perspective. *Science*, 287(5460), March 2000.

[5] Andrew L Hopkins and Colin R Groom. The druggable genome. *Nat Rev Drug Discov*, 1(9), September 2002.

[6] Kenneth A Jacobson and Stefano Costanzi. New insights for drug design from the x-ray crystallographic structures of g-protein-coupled receptors. *Mol. Pharmacol.*, 82(3), September 2012.

[7] Ron O Dror, Morten Jensen, David W Borhani, and David E Shaw. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.*, 135(6), June 2010.

[8] Ron O Dror, Albert C Pan, Daniel H Arlow, David W Borhani, Paul Maragakis, Yibing Shan, Huafeng Xu, and David E Shaw. Pathway and mechanism of drug binding to g-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.*, 108(32), August 2011.

[9] M. J. Harvey, G. Giupponi, and G. De Fabritiis. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.*, 5(6):1632–1639, 2009.

[10] I Buch, M J Harvey, T Giorgino, D P Anderson, and G De Fabritiis. High-throughput all-atom molecular dynamics simulations using distributed computing. *J Chem Inf Model*, 50(3), March 2010.

[11] The universal protein resource (UniProt) in 2010. *Nucleic Acids Res.*, 38(Database issue), January 2010.

[12] Eric Jain, Amos Bairoch, Severine Duvaud, Isabelle Phan, Nicole Redaschi, Baris E Suzek, Maria J Martin, Peter McGarvey, and Elisabeth Gasteiger. Infrastructure for the life sciences: design and implementation of the UniProt website. *BMC Bioinformatics*, 10, 2009.

[13] Elisabeth P Carpenter, Konstantinos Beis, Alexander D Cameron, and So Iwata. Overcoming the challenges of membrane protein crystallography. *Curr Opin Struct Biol*, 18(5), October 2008.

[14] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The protein data bank. *Nucl. Acids Res.*, 28(1):235–242, January 2000.

[15] Gregory L Warren, C Webster Andrews, Anna-Maria Capelli, Brian Clarke, Judith LaLonde, Millard H Lambert, Mika Lindvall, Neysa Nevins, Simon F Semus, Stefan Senger, Giovanna Tedesco, Ian D Wall, James M Woolven, Catherine E Peishoff, and Martha S Head. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 49(20), October 2006.

[16] Hans-Joachim Bohm and Martin Stahl. The use of scoring functions in drug discovery applications. In Kenny B. Lipkowitz and Donald B. Boyd, editors, *Reviews in Computational Chemistry*, pages 41–87. John Wiley & Sons, Inc., 2003.

[17] Brian K Kobilka and Xavier Deupi. Conformational complexity of g-protein-coupled receptors. *Trends Pharmacol. Sci.*, 28(8), August 2007.

[18] Ron O. Dror, Robert M. Dirks, J.P. Grossman, Huafeng Xu, and David E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1), 2012.

[19] New York City. Hpc challenges in computational sciences. June 2013.

[20] Binquan Luan, Martin Caffrey, and Aleksei Aksimentiev. Structure refinement of the OpcA adhesin using molecular dynamics. *Biophys. J.*, 93(9), November 2007.

[21] Karina MartÃnez-Mayorga, Michael C Pitman, Alan Grossfield, Scott E Feller, and Michael F Brown. Retinal counterion switch mechanism in vision evaluated by molecular simulations. *J. Am. Chem. Soc.*, 128(51), December 2006.

[22] Mikhail A Lomize, Andrei L Lomize, Irina D Pogozheva, and Henry I Mosberg. OPM: orientations of proteins in membranes database. *Bioinformatics*, 22(5), March 2006.

[23] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983.

[24] Sunhwan Jo, Joseph B Lim, Jeffery B Klauda, and Wonpil Im. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.*, 97(1), July 2009.

[25] Steven M Foord, Tom I Bonner, Richard R Neubig, Edward M Rosser, Jean-Phillipe Pin, Anthony P Davenport, Michael Spedding, and Anthony J Harmar. International union of pharmacology. XLVI. g protein-coupled receptor list. *Pharmacol. Rev.*, 57(2), June 2005.

[26] Robert Fredriksson, Malin C Lagerstrom, Lars-Gustav Lundin, and Helgi B Schith. The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.*, 63(6), June 2003.

[27] Markus Eilers, Viktor Hornak, Steven O Smith, and James B Konopka. Comparison of class a and d g protein-coupled receptors: common features in structure and activation. *Biochemistry*, 44(25), June 2005.

[28] Vsevolod Katritch, Vadim Cherezov, and Raymond C. Stevens. Diversity and modularity of g protein-coupled receptor structures. *Trends in Pharmacological Sciences*, 33(1):17–27, January 2012.

[29] Matteo Floris, Davide Sabbadin, Ricardo Medda, Alessandro Bulfone, and Stefano Moro. Adenosiland: walking through adenosine receptors landscape. *Eur J Med Chem*, 58, December 2012.

[30] Guillaume Lebon, Tony Warne, Patricia C. Edwards, Kirstie Bennett, Christopher J. Langmead, Andrew G. W. Leslie, and Christopher G. Tate. Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature*, 474(7352):521–525, June 2011.

[31] Fei Xu, Huixian Wu, Vsevolod Katritch, Gye Won Han, Kenneth A. Jacobson, Zhan-Guo Gao, Vadim Cherezov, and Raymond C. Stevens. Structure of an agonist-bound human A2A adenosine receptor. *Science*, 332(6027):322–327, April 2011.

[32] Veli-Pekka Jaakola, Mark T Griffith, Michael A Hanson, Vadim Cherezov, Ellen Y T Chien, J Robert Lane, Adriaan P Ijzerman, and Raymond C Stevens. The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science*, 322(5905), November 2008.

[33] Andrew S DorÃ©, Nathan Robertson, James C Errey, Irene Ng, Kaspar Hollenstein, Ben Tehan, Edward Hurrell, Kirstie Bennett, Miles Congreve, Francesca Magnani, Christopher G Tate, Malcolm Weir, and Fiona H Marshall. Structure of the adenosine a(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure*, 19(9), September 2011.

[34] Tomoya Hino, Takatoshi Arakawa, Hiroko Iwanari, Takami Yurugi-Kobayashi, Chiyo Ikeda-Suno, Yoshiko Nakada-Nakura, Osamu Kusano-Arai, Simone Weyand, Tatsuro Shimamura, Norimichi Nomura, Alexander D. Cameron, Takuya Kobayashi, Takao Hamakubo, So Iwata, and Takeshi Murata. G-protein-coupled receptor inactivation by an allosteric inverse-agonist antibody. *Nature*, 482(7384):237–240, February 2012.

[35] Wei Liu, Eugene Chun, Aaron A Thompson, Pavel Chubukov, Fei Xu, Vsevolod Katritch, Gye Won Han, Christopher B Roth, Laura H Heitman, Adriaan P IJzerman, Vadim Cherezov, and Raymond C Stevens. Structural basis for allosteric regulation of GPCRs by sodium ions. *Science*, 337(6091), July 2012.

[36] Miles Congreve, Stephen P Andrews, Andrew S DorÃ©, Kaspar Hollenstein, Edward Hurrell, Christopher J Langmead, Jonathan S Mason, Irene W Ng, Benjamin Tehan, Andrei Zhukov, Malcolm Weir, and Fiona H Marshall. Discovery of 1,2,4-triazine derivatives as adenosine a(2A) antagonists using structure based drug design. *J. Med. Chem.*, 55(5), March 2012.

[37] David Rodriguez, Xabier Bello, and Hugo Gutierrez-de Teran. Molecular modelling of g protein-coupled receptors through the web. *Molecular Informatics*, 31(5):334–341, 2012.

[38] Bas Vroling, Marijn Sanders, Coos Baakman, Annika Borrmann, Stefan Verhoeven, Jan Klomp, Laerte Oliveira, Jacob de Vlieg, and Gert Vriend. GPCRDB: information system for g protein-coupled receptors. *Nucleic Acids Res.*, 39(Database issue), January 2011.

[39] Catherine L Worth, Annika Kreuchwig, Gunnar Kleinau, and Gerd Krause. GPCR-SSFE: a comprehensive database of g-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics*, 12, 2011.

[40] Yang Zhang, Mark E Devries, and Jeffrey Skolnick. Structure modeling of all identified g protein-coupled receptors in the human genome. *PLoS Comput. Biol.*, 2(2), February 2006.

[41] Ursula Pieper, Benjamin M Webb, David T Barkan, Dina Schneidman-Duhovny, Avner Schlessinger, Hannes Braberg, Zheng Yang, Elaine C Meng, Eric F Pettersen, Conrad C Huang, Ruchira S Datta, Parthasarathy Sampathkumar, Mallur S Madhusudhan, Kimmen SjÃ¶lander, Thomas E Ferrin, Stephen K Burley, and Andrej Sali. ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.*, 39(Database issue), January 2011.

[42] Florian Kiefer, Konstantin Arnold, Michael Konzli, Lorenza Bordoli, and Torsten Schwede. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.*, 37(Database issue), January 2009.

[43] Sebastian Kelm, Jiye Shi, and Charlotte M Deane. MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics*, 26(22), November 2010.

[44] Sunhwan Jo, Taehoon Kim, Vidyashankara G Iyer, and Wonpil Im. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem*, 29(11), August 2008.

[45] Gyrgy Hasko, Joel Linden, Bruce Cronstein, and Pol Pacher. Adenosine receptors: therapeutic aspects for inflammatory and immune diseases. *Nat Rev Drug Discov*, 7(9), September 2008.

[46] C Londos, D M Cooper, and J Wolff. Subclasses of external adenosine receptors. *Proc. Natl. Acad. Sci. U.S.A.*, 77(5), May 1980.

[47] Amotz Rogel, Yael Bromberg, Oded Sperling, and Esther Zoref-Shani. Phospholipase c is involved in the adenosine-activated signal transduction pathway conferring protection against iodoacetic acid-induced injury in primary rat neuronal cultures. *Neurosci. Lett.*, 373(3), January 2005.

[48] Huda E Tawfik, J Schnermann, Peter J Oldenburg, and S Jamal Mustafa. Role of a1 adenosine receptors in regulation of vascular tone. *Am. J. Physiol. Heart Circ. Physiol.*, 288(3), March 2005.

[49] Christa E Muller and Kenneth A Jacobson. Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochim. Biophys. Acta*, 1808(5), May 2011.

[50] B Kull, P Svenningsson, and B B Fredholm. Adenosine a(2A) receptors are colocalized with and activate g(olf) in rat striatum. *Mol. Pharmacol.*, 58(4), October 2000.

[51] Paula Fresco, Carmen Diniz, and Jorge Gonsalves. Facilitation of noradrenaline release by activation of adenosine a(2A) receptors triggers both phospholipase c and adenylate cyclase pathways in rat tail artery. *Cardiovasc. Res.*, 63(4), September 2004.

[52] J W Daly, P Butts-Lamb, and W Padgett. Subclasses of adenosine receptors in the central nervous system: interaction with caffeine and related methylxanthines. *Cell. Mol. Neurobiol.*, 3(1), March 1983.

[53] L E Brackett and J W Daly. Functional characterization of the a2b adenosine receptor in NIH 3T3 fibroblasts. *Biochem. Pharmacol.*, 47(5), March 1994.

[54] Q Y Zhou, C Li, M E Olah, R A Johnson, G L Stiles, and O Civelli. Molecular cloning and characterization of an adenosine receptor: the a3 adenosine receptor. *Proc. Natl. Acad. Sci. U.S.A.*, 89(16), August 1992.

[55] M P Abbracchio, R Brambilla, S Ceruti, H O Kim, D K von Lubitz, K A Jacobson, and F Cattabeni. G protein-dependent activation of phospholipase c by adenosine a3 receptors in rat brain. *Mol. Pharmacol.*, 48(6), December 1995.

[56] Vladimir Shneyvays, Dorit Leshem, Tova Zinman, Liaman K Mamedova, Kenneth A Jacobson, and Asher Shainberg. Role of adenosine a1 and a3 receptors in regulation of cardiomyocyte homeostasis after mitochondrial respiratory chain injury. *Am. J. Physiol. Heart Circ. Physiol.*, 288(6), June 2005.

[57] Martin Englert, Ursula Quitterer, and Karl Norbert Klotz. Effector coupling of stably transfected human a3 adenosine receptors in CHO cells. *Biochem. Pharmacol.*, 64(1), July 2002.

[58] Mayako Michino, Enrique Abola, GPCR Dock 2008 Participants, Charles L. Brooks, J. Scott Dixon, John Moult, and Raymond C. Stevens. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR dock 2008. *Nature Reviews Drug Discovery*, 8(6):455–463, June 2009.

[59] HUAN-XIANG ZHOU and MICHAEL K. GILSON. Theory of free energy and entropy in noncovalent binding. *Chem Rev*, 109(9), September 2009.

[60] John E Ladbury. Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochem. Soc. Trans.*, 38(4), August 2010.

[61] Ernesto Freire. Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today*, 13(19-20), October 2008.

[62] Adam J Ruben, Yoshiaki Kiso, and Ernesto Freire. Overcoming roadblocks in lead optimization: a thermodynamic perspective. *Chem Biol Drug Des*, 67(1), January 2006.

[63] Holger Gohlke and Gerhard Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed. Engl.*, 41(15), August 2002.

[64] R L Rich and D G Myszka. Advances in surface plasmon resonance biosensor analysis. *Curr. Opin. Biotechnol.*, 11(1), February 2000.

[65] P A Kollman, I Massova, C Reyes, B Kuhn, S Huo, L Chong, M Lee, T Lee, Y Duan, W Wang, O Donini, P Cieplak, J Srinivasan, D A Case, and 3rd Cheatham, T E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33(12), December 2000.

[66] Yibing Shan, Eric T Kim, Michael P Eastwood, Ron O Dror, Markus A Seeliger, and David E Shaw. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.*, 133(24), June 2011.

[67] Philip Ball. Water as an active constituent in cell biology. *Chem. Rev.*, 108(1), January 2008.

[68] Andrea Bortolato, Ben G Tehan, Michael S Bodnarchuk, Jonathan W Essex, and Jonathan S Mason. Water network perturbation in ligand binding: Adenosine A2A antagonists as a case study. *J Chem Inf Model*, 53(7), July 2013.