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 Receptors (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 A2A 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 A2A Adenosine Receptor, detailed structural investigation on the dynamic solvation process has been made using state of the art technology 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 accoppiati 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 risultano 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, oppioi e recettori adenosinici, del sottotipo A2A, 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 template, 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 A2A, 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 ortostero dell’recettore adenosinico umano A2A. 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.
List of abbreviations ix
List of original publications xi
Preface xiii

1 Expanding GPCRs structural knowledge through Molecular Modeling 1
  1.1 Introduction to G Protein-Coupled Receptors ............................. 1
    1.1.1 Potential therapeutical application of Adenosine Receptors targeting
    agents .............................................................. 3
  1.2 Exploring the Adenosine Receptors landscape ............................. 4

2 Investigating GPCRs-ligand recognition process 17
  2.1 Modeling the protein-ligand recognition process ....................... 17
  2.2 Methodological advances ................................................. 18
    2.2.1 Molecular Dynamics applicability to Drug Design ............... 19

3 Conclusion and future perspectives 45

Appendix 1: Implementation of the Best Template Searching tool into Adenosiland platform 47

Appendix 2: GPCRs dynamic solvation process: insights using all-atom MD simulations 53

Davide Sabbadin’s publications 89
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.
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:\footnote{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.}


**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.


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.
LIST OF ORIGINAL PUBLICATIONS
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 (A1, A2A, A2B and A3), regulate key processes of the above-mentioned diseases. At the present time, of all the cloned ARs subtypes, only A2A 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.](image_url)

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 A3 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.](image)

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 $\text{Na}^+/\text{Cl}^-$ 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].
1.1 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].
CHAPTER 1. EXPANDING GPCRS STRUCTURAL KNOWLEDGE THROUGH MOLECULAR MODELING

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

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]
1.1. INTRODUCTION TO G PROTEIN-COUPLED RECEPTORS

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<sub>2A</sub> 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<sub>2A</sub> 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<sub>2A</sub> receptor/T4 lysozyme chimera (PDB ID: 3EML [32]) and with other hA<sub>2A</sub> AR mutants/chimeras (PDB ID: 3PWH [33] PDB ID: 3VGA [34], PDB ID: 3VG9 [34] and PDB ID: 4E1Y [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<sub>2A</sub> 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<sub>1</sub> AR subtype mediates an inhibition of adenylyl cyclase through activation of pertussis toxin-sensitive G<sub>i/o</sub> proteins [46] and results in increased activity of PLC [47] [48]. High levels of A<sub>1</sub> Adenosine Receptor expression were found in the brain, heart,
adipose tissue, stomach, testis, spleen, kidney, aorta, liver, eye and bladder [45]. Receptor 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, Huntington’s 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.
Adenosiland: Walking through adenosine receptors landscape

Matteo Floris, Davide Sabbadi, Ricardo Medda, Alessandro Bulfone, Stefano Moro

Adenosine receptors (ARs) belong to the family of G protein-coupled receptors. Four distinct subtypes are known, termed adenosine A1, A2A, A2B and A3 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 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/.

© 2012 Elsevier Masson SAS. All rights reserved.

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 A1, A2A, A2B and A3 [3]. These receptors are generally coupled to adenylate cyclase. Activation of the A1 and A3 receptors has an inhibitory effect, whereas A2A and A2B 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 receptors subtypes which belong to the superfamilly 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 (A1 and A3 subtypes) or activation, via G_{o} protein (A2A and A2B 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 A2A receptor in complex with different agonists and antagonists have been solved and released. The ARs physiological
agonist adenosine (PDB ID: 2YDO), its N-ethyl-S-carboxamide derivative, NECA (PDB ID: 2YDV [8]) and the high affinity agonist UK-432957, 6-(2,2-diphenylethylamino)-9-(2R,3R,4S,5S)-3-(ethyl carbamoyl)-3,4-dihydroxy-oxolan-2-yl)-N-[2-((1-pyridin-2-ylpyperidin-4-yl)carbamoylamino)ethyl]purine-2-carboxamide (PDB ID: 3QAK [9]) have been co-crystallized with the human A2A AR. Moreover, antagonists belonging to different chemical families, have been also co-crystallized with the human A2A AR. In particular, the high affinity antagonist (4-2-[7-amino-2-(2-furyl)] [1,2,4]triazolo[2,3-z] [1,3,5]triazin-5-ylamino)ethyl)phenol, better known as ZM241385, is in complex with the human Adenosine A2A receptor/T4 lysosome chimera (PDB ID: 3EML [10]) and in complex with other hA2A AR mutants/chimeras (PDB ID: 3PWH [11] PDB ID: 3VGA [12], PDB ID: 3VGG [12] and PDB ID: 4EY [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)phenoxo)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 A2A 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]; Javalview, a Java Alignment Editor [20,21]; Yasara modeling suite (version 11.4.18) [22]; Molecular Operating Environment suite (MOE[23]; MEGA 5.0 [24] and RDa, 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 Erhl for his courtesy in using its the Java Molecular 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 A2A receptor bound to the high affinity antagonist ZM241385 (PDB ID: 3EML) [10] has been used as template structure. The lysosome 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 A2B receptor models.

- All human A1, A2A, A2B and A3 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 A2A receptors available at the time of the preparation of this
manuscript (in total 12 structures: PDB ID: 2YDO, 2YDV [8], 3EM1 [10], 3QAK [9], 3PWH, 3REY, 3RFM [11], 3UZA, 3ZUC [14], 3VGA, 3VC9 [12], 4EY [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/j and χ1/χ2 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 Berendsen 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 CRST 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 A1 receptors, six A2A receptors, seven A2B receptors, and seven A3. 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], CHARMM-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 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.
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 a 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 external agonist/antagonist and all co-crystallized adenosine ligands. A practical download bar is located at the bottom of Adenosiland 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 a 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 · pharmacophoric similarity + 0.4 · 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
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 A2A subtype is characterized by a longer primary sequence mainly due to its C-terminal tail (>110 amino acids). It is very well documented that the C-terminus of adenosine A2A 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 A2A 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, A2B subtypes are characterized by the longest ECL2 (>38 amino acids) where, in A3 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 A2A subtype. The two cysteines involved in the formation of this additional ECL2-ECL1 disulfide bridge are conserved in all analyzed A2A receptors. Additionally, the A2A subtype also possesses 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 A2B receptor models, is detectable [40]. This information has been taken into consideration during the homology model building procedure of all A2B receptors. A summary of conserved cysteines putatively involved 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, A3 receptors from sheep (Ovis aries P35342), dog (Canis familiaris Q28309) and mouse (Mus musculus Q61618) share a potential glycosylation sites on ECL2. Besides, A2A and A2B, A1 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 A2A 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 A3 receptor subtypes and in the A2A receptor of guinea pig [65]. Beside these two...
mutations at the 6.48 and 6.52 positions that apparently can impair receptor activity, the A2A 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 A3 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 A2A receptor crystal structures highlights that the differences in geometrical positions of alpha-carbons of the orthosteric binding 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 A3 adenosine receptor models generating by all possible template all crystallized A2A 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 A2A adenosine receptor X-ray structures show a volumetric range from ca. 1970 Å³ (PDB ID: 2YDO) to ca. 1120 Å³ (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...
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 bio-informatics 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>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison of the most relevant bioinformatics information regarding all adenosine receptor subtypes.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniprot ID</td>
<td>Seq. length (AAs)</td>
</tr>
<tr>
<td>A1</td>
<td>Dog P11616</td>
</tr>
<tr>
<td>Rat P25099</td>
<td>326</td>
</tr>
<tr>
<td>Bovine P28190</td>
<td>326</td>
</tr>
<tr>
<td>Human P30542</td>
<td>326</td>
</tr>
<tr>
<td>Rabbit P34970</td>
<td>328</td>
</tr>
<tr>
<td>Guinea Pig P47745</td>
<td>326</td>
</tr>
<tr>
<td>Chicken P49882</td>
<td>324</td>
</tr>
<tr>
<td>Sumatra orangutan Q5RF57</td>
<td>326</td>
</tr>
<tr>
<td>Mouse Q60612</td>
<td>326</td>
</tr>
<tr>
<td>A2A</td>
<td>Human P29274</td>
</tr>
<tr>
<td>Dog P11617</td>
<td>412</td>
</tr>
<tr>
<td>Rat P30543</td>
<td>410</td>
</tr>
<tr>
<td>Guinea Pig P46616</td>
<td>409</td>
</tr>
<tr>
<td>Mouse Q60613</td>
<td>410</td>
</tr>
<tr>
<td>Horse Q6TL17</td>
<td>412</td>
</tr>
<tr>
<td>A2B</td>
<td>Chicken P29275</td>
</tr>
<tr>
<td>Human P29275</td>
<td>332</td>
</tr>
<tr>
<td>Rat P29276</td>
<td>332</td>
</tr>
<tr>
<td>Bovine P11617</td>
<td>332</td>
</tr>
<tr>
<td>Rabbit P34970</td>
<td>332</td>
</tr>
<tr>
<td>Mouse Q60614</td>
<td>332</td>
</tr>
<tr>
<td>Dog Q6W3F4</td>
<td>332</td>
</tr>
<tr>
<td>A3</td>
<td>Rabbit O02667</td>
</tr>
<tr>
<td>Rat P28647</td>
<td>320</td>
</tr>
<tr>
<td>Human P33765</td>
<td>318</td>
</tr>
<tr>
<td>Sheep P35342</td>
<td>317</td>
</tr>
<tr>
<td>Bovine Q0VC81</td>
<td>317</td>
</tr>
<tr>
<td>Dog Q28309</td>
<td>314</td>
</tr>
<tr>
<td>Mouse Q61618</td>
<td>319</td>
</tr>
</tbody>
</table>

*a* % of similarities and % of identities are calculated using the human A2A primary sequence as reference sequence (UniProt ID: P29274).

*a* Root mean square deviations (r.m.s.d., in Å) are calculated using the crystallographic structure 3EML [10] as a reference structure.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Disulfide bond mapping of adenosine receptor subtypes. All modeled receptors have been constructed following disulfide bridges information reported in the present table.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receptor subtype</td>
<td>TM3-ECL2 (C77-C166)*</td>
</tr>
<tr>
<td>A1</td>
<td>Yes</td>
</tr>
<tr>
<td>A2A</td>
<td>Yes</td>
</tr>
<tr>
<td>A2B</td>
<td>Yes</td>
</tr>
<tr>
<td>A3</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*a* Numbering is referred to the human A2A 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 extending the similarity search screening to BindingDB. It provides 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 BIND [71].

Acknowledgment

The work was supported by a grant of the Italian Ministry for University and Research (MIUR, FIRB RBNE03YAL3 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 Acelera 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


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 to accurately reproduce binding affinity values that are consistent with ITC [64]. From a computational standpoint, numerous successful attempts have 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 ($\Delta G$), 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.](image)

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.
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. They regulate several crucial functions of most cells in the body, and receptor dysfunction can lead to a variety of disease conditions. 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. 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. It is estimated that these receptors represent about one-third of the actual identified targets of clinically used drugs. 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.

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. 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. 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. 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 have 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
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. 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. 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. 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.

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.

To validate our combined approach, we have selected, as a key study, the human A2A adenosine receptor (hA2A 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 hA2A 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. Recently, phase III studies on Preladenant, which is an A2A 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.

Figure 1. Overview of the binding modes at the hA2A AR of the selected co-crystallized ligands: (A) ZM 241385 (PDB ID: 3EML), (B) T4E (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.
For the present study, we have selected five crystal structures of the hA2A AR in complex with four strong binders, such as 4-[(2-amino-3-carboxamido)adenosine (pK_D = 9.18 ± 0.00, PDB ID: 3EM1)], 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pK_D = 8.9 ± n.d., PDB ID: 3UZA), 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pK_D = 9.6 ± n.d., PDB ID: 3UZC); NECA, 3H-5-carboxamido adenosine (pK_D = 7.00 ± 0.1, PDB ID: 2YDV) and a weaker binder such as caffeine (pK_D = 5.31 ± 0.44, PDB ID: 3RFM). The structures of the considered antagonists inside the orthosteric binding pocket of the hA2A 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 4 NVIDIA GTX 580 and 2 NVIDIA GTX 680 GPU cluster engineered by Accelrys. 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.

### Homology Models.

The selected five crystal structures (PDB IDs: 3EM1, 3UZA, 3UZC, 3RFM, and 2YDV) and the FASTA sequence of the hA2A AR (UniProt ID: P29274) were retrieved from the RCSB PDB database (http://www.rcsb.org) and the UniProtKB/Swiss-Prot, respectively. The eventual lysosome 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 tool36 and overlapping lipids (within 0.6 Å) were removed using a grid-based method35 with the VMD Membrane Plugin. 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_elec and IE_hydro, 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,i.e.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.6 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.

### 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, 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 hA2A AR models with the GOLD 5.1 suite using the genetic algorithm protocol (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. The latter, in a previous study, resulted the best among the tested scoring functions in reproducing and ranking the crystallographic binding mode of ZM 241385 at the hA2A AR. 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_elec and IE_hydro, 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,i.e., 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. 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 for the hA2A AR in complex with the antagonist T4G (PDB ID: 3UZA). The membrane has been generated by using a grid-based method with the VMD Membrane Plugin tool 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
representation is depicted in Figure S2 in the Supporting Information (panel 1).

The prepared systems were solvated with TIP3P37 water using the Solvate 1.0 program38 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,39 using the CHARMM27 Force Field40 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 theory42 (which is consistent with the CHARMM27 Force Field parameterization) using Gaussian0935 and the imple-
mented parameterization 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 Berendsen barostat. Bond lengths involving hydrogen atoms were constrained using the M-SHAKE44 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 were 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 (IEele) and hydrophobic (IEhyd) 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 (DSFele, eq 1) and the hydrophobic DSF (DSFhyd, eq 2):

\[
\text{DSFele} = \sum_{i=0}^{n} \text{IE}_{\text{ele}}
\]

\[
\text{DSF}_{\text{hyd}} = \sum_{i=0}^{n} \text{IE}_{\text{hyd}}
\]

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 wDSFele and wDSFhyd respectively) are reported in eqs 3 and 4:

\[
\text{wDSF}_{\text{ele}} = \frac{\sum_{i=0}^{n} \text{IE}_{\text{ele}}}{\text{RMSD}}
\]

\[
\text{wDSF}_{\text{hyd}} = \frac{\sum_{i=0}^{n} \text{IE}_{\text{hyd}}}{\text{RMSD}}
\]

The obtained DSF and wDSF values then were plotted against the simulation time and generic linear functions (f(x) = m·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,48 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 hA2A AR. The binding site of the hA2A AR has been exhaustively described elsewhere. 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 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),49 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
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 data48 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.

Work flow 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 hA2A 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).

Figure 2. Workflow of the combined molecular docking and membrane molecular dynamics protocol.
In Silico Inspection of ZM 241385 Binding Mode. We first tested our combined procedure with the potent hA2A 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 (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...
Figure 4. (A) Docking poses of T4G at the hA2A 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.
Figure 5. (A) Docking poses of T4E at the hA2A 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.
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 ($5.33 \pm 1.87 \AA$) 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 $2.47 \AA$, 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 hA2A AR, an aspect which has been extensively and uniquely described for the $\beta_2$-adrenergic receptor. 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 hA2A 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 S-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 = $0.99 \pm 0.43 \AA$). 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: $2.63 \pm 0.96 \AA$; pose 2: $1.78 \pm 0.53 \AA$).

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.,
Figure 6. (A) Docking poses of caffeine at the hA2A 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 hA3 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 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

Figure 7. (A) Docking poses of NECA at the hA2A AR with their corresponding (B) static IEFs and (C) dynamic IEFs, and (D) electrostatic and (E) hydrophobic wDSFs. IEFH values are given in kcal Å−1 mol−1, IEFH values are given in arbitrary hydrophobic units, and ligand fluctuations (average RMSD reported on top of the IEFs) are given in Å.
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 hA2A 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 π−π stacking interaction with the conserved Phe168 (EL2) side chain. The ribose moiety, which is deeply inserted in the orthosteric binding pocket and plays a crucial role in antagonist or agonist recognition, in the hA2A 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, a proposed analysis of dynamic IEFs, ligand positional stability (RMSD profile), and DSF 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 hA2A 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 hA2A 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 hA2A 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 and recently clarified for the hA2A AR. 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 permanence of water molecules within 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), 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, but that has not been yet reported for ZM 241385, although, in recent X-ray structures, 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), 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
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 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 an explicit lipid–water environment.

We selected, as a test case, the human A2A adenosine receptor (hA2A AR) in complex with four antagonists—namely, ZM 241385, T4G, T4E, and caffeine—and one agonist (N-ethyl-S-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) 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-S-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**


(8) Böhm, H.-J.; Stahl, M. The Use of Scoring Functions in Drug Discovery Applications. In *Reviews in Computational Chemistry*;


(57) Kufareva, I.; Katritch, 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.
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. 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. 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. 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 pharmacodynamic 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. 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 enables 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 (dcms) is monitored over a fixed time window (∆t= 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 \( f(x) = mx \) 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 (dcms) is less than 5 Å. To validate the methodology, we selected as a key study the human A2A Adenosine Receptor (hA2A AR), that has been recently crystalized with several ligands, both agonists and antagonists, characterized by different receptor binding affinities. In particular, we selected four crystal structures of the hA2A 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, 2M 241385 (pKₐ = 9.18 ± 0.00); 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pKₐ = 8.9 ± n.d.); PDB code: 3EML; 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pKₐ = 8.9 ± n.d.); PDB code: 3UZA;
4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pK_D = 9.6 ± n.d., PDB code: 3UZC) ; and a weaker binder such as caffeine (pK_D = 5.31 ± 0.44, PDB code: 3RFM).

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 exhibits 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 A2A Adenosine Receptor, the structural information extracted from SuMD simulation is indistinguishable (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 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 and in some cases they may coincide with possible allosteric sites. 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 A2A 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 A2A 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.

Figure 2. Panel A: Overview of the Adenosine Receptor Antagonist ZM241385-human A2A 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 A2A 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 hA2A 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).
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 A2A 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 A2A 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 (dcmL-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 A2A 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 A2A Adenosine Receptor recognition mechanism.

The Extracellular Loop 3 (EL3) of the human A2A 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 A2A 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 (dcmL-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 A2A 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 A2A Adenosine Receptor recognition mechanism.

As reported for the high-affinity human A2A 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 A2A Adenosine Receptor recognition mechanism using Supervised Molecular Dynamics (SuMD). Ligand-Receptor distance vector (\(d_{cm, L-R}\)) 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 A2A 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. 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 (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.
Supporting Information. Complete experimental section and additional result discussion for ZM241385, T4E, T4G and caffeine-human A3AR 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; ZM241385: 4-(6-(7-amino-2-(2-furyl)(1,2,4)triazolo(2,3-a)1(3,5)triazin-5-yl-amino)ethyl)phenol.

REFERENCES


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

Implementing the “Best Template Searching” tool into Adenosiland platform

Matteo Floris¹, Davide Sabbadin², Antonella Ciancetta², Ricardo Medda¹, Alberto Cuzzolin² 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 A2A AR (h A2A 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 A1, A2B, and A3 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 hA2A 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 GCPRs.

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 hA2A 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 A1, A2A, A2B and A3 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 A2A 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
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 hA2A 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 and 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 A1, 469 A2A, 559 A2B and 290 A3 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:

<table>
<thead>
<tr>
<th>Input ligand</th>
<th>Suggested template</th>
<th>Combined similarity value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine</td>
<td>2YDO</td>
<td>0.83</td>
</tr>
<tr>
<td>NECA</td>
<td>2YDO</td>
<td>0.72</td>
</tr>
<tr>
<td>UK-432097</td>
<td>3QAK</td>
<td>0.37</td>
</tr>
<tr>
<td>ZMA 241385</td>
<td>4EIY</td>
<td>0.69</td>
</tr>
<tr>
<td>T4G</td>
<td>3UZA</td>
<td>0.84</td>
</tr>
<tr>
<td>T4E</td>
<td>3UZC</td>
<td>0.92</td>
</tr>
<tr>
<td>XAC</td>
<td>3REY</td>
<td>0.67</td>
</tr>
<tr>
<td>Caffeine</td>
<td>3RFM</td>
<td>0.98</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Input ligand</th>
<th>Suggested template</th>
<th>Combined similarity value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenosine</td>
<td>2YDO</td>
<td>0.83</td>
</tr>
<tr>
<td>NECA</td>
<td>2YDO</td>
<td>0.72</td>
</tr>
<tr>
<td>UK-432097</td>
<td>3QAK</td>
<td>0.37</td>
</tr>
<tr>
<td>ZMA 241385</td>
<td>4EIY</td>
<td>0.69</td>
</tr>
<tr>
<td>T4G</td>
<td>3UZA</td>
<td>0.84</td>
</tr>
<tr>
<td>T4E</td>
<td>3UZC</td>
<td>0.92</td>
</tr>
<tr>
<td>XAC</td>
<td>3REY</td>
<td>0.67</td>
</tr>
<tr>
<td>Caffeine</td>
<td>3RFM</td>
<td>0.98</td>
</tr>
</tbody>
</table>

### Table 2 Similarity sorting of human A2A AR templates based on furan-2-carboxylic acid (4-phenyl-5-pyridin-4-yl-thiazol-2-yl)-amide query ligand

<table>
<thead>
<tr>
<th>Ligand</th>
<th>PDB ID template</th>
<th>Shape similarity</th>
<th>2D similarity (Tanimoto)</th>
<th>2D similarity (Tversky)</th>
<th>Pharmacophore similarity (Tanimoto)</th>
<th>Pharmacophore similarity (Tversky)</th>
<th>Combined similarity (Shape &amp; FP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4G</td>
<td>3UZA</td>
<td>0.33</td>
<td>0.86</td>
<td>0.89</td>
<td>0.46</td>
<td>0.65</td>
<td>0.52</td>
</tr>
<tr>
<td>ZM 241385</td>
<td>3PWH</td>
<td>0.58</td>
<td>0.90</td>
<td>0.93</td>
<td>0.27</td>
<td>0.42</td>
<td>0.48</td>
</tr>
<tr>
<td>T4E</td>
<td>3UZC</td>
<td>0.37</td>
<td>0.84</td>
<td>0.89</td>
<td>0.44</td>
<td>0.54</td>
<td>0.47</td>
</tr>
<tr>
<td>ZM 241385</td>
<td>4EIY</td>
<td>0.34</td>
<td>0.90</td>
<td>0.93</td>
<td>0.27</td>
<td>0.43</td>
<td>0.39</td>
</tr>
<tr>
<td>ZM 241385</td>
<td>3EML</td>
<td>0.35</td>
<td>0.90</td>
<td>0.93</td>
<td>0.27</td>
<td>0.42</td>
<td>0.39</td>
</tr>
<tr>
<td>NECA</td>
<td>2YDV</td>
<td>0.51</td>
<td>0.82</td>
<td>0.87</td>
<td>0.17</td>
<td>0.31</td>
<td>0.39</td>
</tr>
<tr>
<td>ZM 241385</td>
<td>3VG9</td>
<td>0.32</td>
<td>0.90</td>
<td>0.93</td>
<td>0.27</td>
<td>0.43</td>
<td>0.38</td>
</tr>
<tr>
<td>XAC</td>
<td>3REY</td>
<td>0.21</td>
<td>0.89</td>
<td>0.94</td>
<td>0.25</td>
<td>0.48</td>
<td>0.37</td>
</tr>
<tr>
<td>ZM 241385</td>
<td>3VGA</td>
<td>0.28</td>
<td>0.90</td>
<td>0.93</td>
<td>0.27</td>
<td>0.42</td>
<td>0.36</td>
</tr>
<tr>
<td>Adenosine</td>
<td>2YDO</td>
<td>0.33</td>
<td>0.82</td>
<td>0.86</td>
<td>0.18</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Caffeine</td>
<td>3RFM</td>
<td>0.26</td>
<td>0.81</td>
<td>0.85</td>
<td>0.21</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>UK-432097</td>
<td>3QAK</td>
<td>0.16</td>
<td>0.87</td>
<td>0.93</td>
<td>0.14</td>
<td>0.35</td>
<td>0.27</td>
</tr>
</tbody>
</table>
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 A1, A2B and A3 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 hA2A 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]acetonitrile.

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

1CRS4, Parco Polaris, 09010 Pula, CA, Italy. 2Molecular 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
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<sub>2A</sub> antagonists using structure based
Doré AS, Robertson N, Erey JC, Ng I, Hollenstein K, Tehan B, Hurrell E, Bennett K,
adenosine A<sub>2A</sub> receptor in complex with ZM241385 and the xanthines XAC
and caffeine. Structure 19:1283–1293
chemical space using pepMMIMIC. Nucleic Acids Res 39:W261–W269
through adenosine receptors landscape. Eur J Med Chem 58:248–257
Union of Pharmacology. XXV. Nomenclature and classification of adenosine
receptors. Pharmacol Rev 53:527–552
Gauthon A, Bellis LJ, Berto AP, Chambers J, Davies M, Hersey A, Light Y,
large-scale bioactivity database for drug discovery. Nucleic Acids Res
40:D1100–D1107
Hino T, Arakawa T, Iwamori H, Yurugi-Kobayashi T, Ikeda-Suno C, Nakada-Nakura Y,
Inamdar GS, Pandya AN, Thakar HM, Sudarsanam V, Kechler S, Sabbadin D, Moro
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, Cheerezov V, Chien EYT, Lane JR, Ijzerman AP,
Stevens RC (2008) The 2.6 angstrom crystal structure of a human A<sub>2A</sub>
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<sub>2A</sub> receptor structures reveal common
Liu W, Chun E, Thompson AA, Chubukov P, Xu F, Katritch V, Han GW, Roth CB,
Heitman LH, Ijzerman AP, Cheerezov V, Stevens RC (2012) Structural basis for
allosteric regulation of GPCRs by sodium ions. Science 337:232–236
chemistry development kit (CDK): an open-source Java library for
developments of the chemistry development kit (CDK) an open-source java
Xu F, Wu H, Katritch V, Han GW, Jacobson KA, Gao Z-G, Cheerezov V, Stevens RC
(2011) Structure of an agonist-bound human A<sub>2A</sub> adenosine receptor.
Science 332:322–327

Cite this article as: Floris et al.: Implementing the “Best Template
Searching” tool into Adenosiland platform. In Silico Pharmacology
2013 1:25.

Submit your manuscript to a SpringerOpen journal and benefit from:
► Convenient online submission
► Rigorous peer review
► Immediate publication on acceptance
► Open access: articles freely available online
► High visibility within the field
► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
Appendix 2: GPCRs dynamic solvation process: insights using all-atom MD simulations


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.
Perturbation of water's fluid dynamics properties during GPCR-ligand recognition: the human A$_2$A 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-oleyl-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.

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.

To validate our key study approach, we have selected the human A2A 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 hA2A 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 (pKD = 9.18 ± 0.00, PDB ID: 3EML); 6-(2,6-dimethylpyridin-4-yl)-5-phenyl-1,2,4-triazin-3-amine, T4G (pKD = 8.9 ± n.d., PDB ID: 3UZA); 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol, T4E (pKD = 9.6 ± n.d., PDB ID: 3UZC); NECA: N-Ethyl-5'-Carboxamido Adenosine (pKD = 7.00 ± 0.1, PDB ID: 2YDV) and a weaker binder such as caffeine (pKD = 5.31 ± 0.44, PDB ID: 3RFM). A detailed inspection of the caffeine binding mode and hydro-dynamics is reported in the Supporting Information section. In addition to the exploration of the hydrodynamic profile of most of the crystallographic structures of A2A AR available at on the RCSB Protein Data Bank, in order to gather insights on the perturbation of
water's fluid dynamics properties during hA\(_2\A\)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\(_2\A\)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\(_2\A\)AR antagonists, considered in this study, have been recently synthesized and reported in literature\(^{15}\).

Table 1

<table>
<thead>
<tr>
<th>Compound</th>
<th>SlogP(o/w)</th>
<th>Vdw volume (Å)</th>
<th>MW (Da)</th>
<th>pKi</th>
<th>kD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,2,4-triazin-3-amine</td>
<td>-0.34</td>
<td>115.50</td>
<td>96.093</td>
<td>ND</td>
</tr>
<tr>
<td>2 (4a)(^{15})</td>
<td>diphenyl-1,2,4-triazine-3-amine</td>
<td>3.49</td>
<td>345.45</td>
<td>248.28</td>
<td>6.93</td>
</tr>
<tr>
<td>3 (4d)(^{15})</td>
<td>6-(3,5-dimethylphenyl)-5-phenyl-1,2,4-triazin-3-amine</td>
<td>4.20</td>
<td>394.30</td>
<td>276.34</td>
<td>7.67</td>
</tr>
<tr>
<td>4</td>
<td>5-phenyl-6-(pyridin-4-yl)-1,2,4-triazin-3-amine</td>
<td>2.26</td>
<td>336.37</td>
<td>249.27</td>
<td>ND</td>
</tr>
<tr>
<td>5 (4g)(^{15})</td>
<td>PDB ID: 3UZA</td>
<td>2.86</td>
<td>385.23</td>
<td>277.33</td>
<td>8.11</td>
</tr>
</tbody>
</table>
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\textsuperscript{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\textsuperscript{19}.

Homology models. The selected five crystal structures (PDB IDs: 3EML, 3UZA, 3UZC, 3RFM and 2YDV) and the FASTA sequence of the hA\textsubscript{2A} AR (Uniprot ID: P29274) were retrieved from the RCSB PDB database\textsuperscript{17} (http://www.rcsb.org) and the UniProtKB/Swiss-Prot\textsuperscript{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\textsuperscript{23}. Then, to minimize contacts among hydrogens, the structures were subjected to energy minimization with Amber99 force field\textsuperscript{24} until the \textit{r.m.s.} of conjugate gradient was $< 0.05 \text{ kcal-mol}^{-1}.\text{\AA}^{-1}$, 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 \textit{r.m.s.} of conjugate gradient $< 0.05 \text{ kcal-mol}^{-1}.\text{\AA}^{-1}$ was reached. Missing loop domains were constructed by the loop search method implemented in Molecular Operating Environment (MOE, version 2012.10) program\textsuperscript{25} on the basis of the structure of compatible fragments found in the Protein Data Bank\textsuperscript{17}. 
N-terminal and C-terminal were deleted if their lengths exceeded those found in the crystallographic template. The “Protonate-3D” tool\textsuperscript{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 $\chi$ plots measure $\phi/\psi$ and $\chi_1/\chi_2$ 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\textsuperscript{26} for the hA$_{2\alpha}$ AR in complex with the antagonist T4G (PDB ID: 3UZA\textsuperscript{15}). Ligand-receptor complexes, where the crystal structure was not available, were obtained by molecular docking using the protocol described previously\textsuperscript{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\textsuperscript{27} water using the program Solvate 1.0\textsuperscript{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\textsuperscript{12} using the CHARMM27 Force Field\textsuperscript{29} and periodic boundaries conditions. Initial parameters for the ligands were derived from the CHARMM General Force Field for organic molecules\textsuperscript{30} by using the “paramchem” service\textsuperscript{31,32} and were subsequently optimized at the MP2/6-31G* level of theory\textsuperscript{33} (consistently with the CHARMM27 Force Field parameterization) by using Gaussian 09\textsuperscript{34} and the implemented parameterization tools in the VMD engine\textsuperscript{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·Å² 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·Å². 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-SHAKE³⁶ 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 VMD³⁵, WORDOM³⁸, the PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC., Gnuplot graphic utility (http://www.gnuplot.info/) and Gromacs tools³⁹.

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. 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 A2A 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 hA2A AR.

The binding site of the hA2A AR has been exhaustively described elsewhere. 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 A2A 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], 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, showing loss of affinity for the Asn253 [6.55] mutant, as well as with recent mutagenesis data 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 A2A AR 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 A2A 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 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.

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$^{-1}$.

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.

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), its N-ethyl-5′carboxamide derivative, NECA,
(PDB ID: 2YDV\textsuperscript{[16]}) and the high affinity agonist UK-432097, 6-(2,2-diphenylethlamino)-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\textsuperscript{[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]-triazol[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\textsuperscript{[14]}) and in complex with other hA\textsubscript{$A_{2A}$} AR mutants/chimeras (PDB ID: 3PWH\textsuperscript{[13]} PDB ID: 3VGA\textsuperscript{[55]}, PDB ID: 3VG9\textsuperscript{[55]} and PDB ID: 4EI\textsuperscript{[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\textsuperscript{[13]}) and the very well known caffeine (PDB ID: 3RFM\textsuperscript{[13]}) have been co-crystallized with A\textsubscript{$A_{2A}$} 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\textsuperscript{[15]}) and the 4-(3-amino-5-phenyl-1,2,4-triazin-6-yl)-2-chlorophenol (PDB ID: 3UZC\textsuperscript{[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\textsubscript{$A_{2A}$} AR revealed the role of water molecules in mediating protein-ligand binding.

The heat-maps corresponding to the ZM241385-hA\textsubscript{$A_{2A}$} 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 shown in figure 3b. In fact, the arrangement of crystallized water molecules in the crystal structure of the chimeric protein of A\textsubscript{2A} AR-BRIL in complex with ZM241385 at 1.8Å resolution (PDB ID: 4EIY\textsuperscript{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\textsubscript{2A} AR in complex with the antagonist ZM241385 (PDB ID: 3EML\textsuperscript{14} – figure 3c; PDB ID: 3VG9\textsuperscript{55} – figure 3d) reveal a similar solvation pattern but in a lower level of detail.

In the agonist bound (NECA) hA\textsubscript{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\textsubscript{2A} AR with NECA bound (PDB ID: 2YDV\textsuperscript{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\textsubscript{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\textsubscript{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 (\textbf{compound 2, table 1}) to 345.45 Å\textsuperscript{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 (\textbf{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 (\textbf{wB} cluster, \textbf{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 \textbf{compounds 2} and \textbf{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 (\textbf{compound 4, table 1}) results in the creation of two “non-bulk” structural water molecules arrangements (\textbf{wA} and \textbf{wB} clusters, \textbf{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 hA2A 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)\textsuperscript{15} of compound 5 (1.15×10\textsuperscript{-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


(33) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. 

(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.; 
Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, 
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.; 
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.


(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 


(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 
1718.

*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 

(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 

chains in the third and seventh transmembrane helical domains of human A2A adenosine 

(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 

(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 


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-hA2A 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 hA2A 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.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.


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


[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.


