



TESIS DOCTORAL

**Computational Approaches for the Characterization of the
Structure and Dynamics of G Protein-Coupled Receptors:
Applications to Drug Design**

David Rodríguez Díaz

Departamento de Anatomía Patolóxica e Ciencias Forenses

Facultade de Medicina e Odontoloxía

Universidade de Santiago de Compostela



DEPARTAMENTO DE ANATOMÍA PATOLÓXICA
E CIENCIAS FORENSES

Facultade de Medicina e Odontoloxía
R/ de San Francisco, s/n
15782 - Santiago de Compostela
Teléfono: 881812215. Fax: 981580336

D. Ángel Carracedo Álvarez, profesor catedrático do Departamento de Anatomía Patolóxica e Ciencias Forenses en calidade de Titor; e **D. Hugo Gutiérrez de Terán** Castañón, doutor e investigador Isidro Parga Pondal da Fundación Pública Galega de Medicina Xenómica (SERGAS) como Director da Tese de Doutoramento realizada por **D. David Rodríguez Díaz** titulada "Computational Approaches for the Characterization of the Structure and Dynamics of G Protein-Coupled Receptors: Applications to Drug Design".

FAN CONSTAR:

Que o citado traballo de investigación reúne os requisitos académicos e científicos necesarios para proceder á súa lectura e defensa pública.

E para que así conste, damos o visto e prace en Santiago de Compostela a 12 de setembro de 2012.

Prof. Ángel Carracedo Álvarez
Titor da Tese de Doutoramento

Dr. Hugo Gutiérrez de Terán
Director da Tese de Doutoramento

D. David Rodríguez Díaz
Autor da Tese de Doutoramento

La realización de esta tesis doctoral ha sido financiada por el Fondo de Investigación en Salud (Instituto Carlos III) mediante la Ayuda Predoctoral con expediente FI08/0799.

Adicionalmente, trabajos aquí presentados se enmarcan dentro de Proyectos de Investigación financiados por la Consellería de Sanidade, Xunta de Galicia (PS09/63) y del Ministerio de Ciencia e Innovación (SAF2011-30104).

Finalmente, se reconoce el tiempo de cómputo y asistencia ofrecida por el Centro de Supercomputación de Galicia (CESGA), en el cual se ejecutaron dos proyectos ICTS (MICINN).



XUNTA DE GALICIA
CONSELLERÍA DE SANIDADE



A Tamara

AGRADECIMIENTOS

Por desgracia, esta sección de agradecimientos no recogerá todas las contribuciones que multitud de personas han realizado para que llegue este momento. Independientemente de su magnitud o naturaleza, dichas aportaciones cobran un especial valor echando la vista atrás, contradiciendo la sombría imagen que se atribuye a la sociedad de hoy en día.

Gracias a tantos amigos de As Pontes de los que he podido disfrutar de su complicidad y compañía, a pesar de que no entendáis muy bien en lo que trabajo ;) Basoa, Juanín, Fer, Rozenn (gracias por la genial portada artista!) y otros tantos, sois grandes! Por supuesto, lo mismo es aplicable a los compañeros de carrera, tanto en la USC como en la UAM, que hicieron que aquellos años fuesen maravillosos y siguen ahí para lo que haga falta: Jaime, Nuria, Álex; Lorena, Erika, Eli, Abel, las Rocíos... Igualmente hay que agradecer el apoyo y consejos de grandes profesores, especialmente a Cristina Murga.

En el ámbito más profesional, un recuerdo para José Antonio, Julián y Alejandro (UDC) en la primera etapa en la Rede Galega de Bioinformática, además de a otros miembros de ese rico colectivo que esperemos continúe vivo de una u otra forma. Asimismo, gracias a Eddy y Vicen por todo el valor que habéis dado a nuestro trabajo computacional y por vuestra cercanía, ojalá podamos continuar por esta senda! También gracias a José Manuel y Pepo de Biofarma por vuestra amabilidad y disponibilidad, al igual que a Mabel Loza por su accesibilidad y consejos. Igualmente gracias a tantos científicos con los que he podido coincidir en distintos eventos y colaboraciones, todas sus aportaciones son significativas.

Of course thanks to Andreas Bender, for being such a great host in Cambridge, together with Pete and Bobby. It has been an incredible experience to be in your laboratory for a while, for providing a stimulant scientific environment and a human group impossible to match. A big hug to all my fellows there: Giorgos, Eva, Shardul, Aakash, Tere, Pedro, Maite, Sonia, Florian, Susana, Richard, Hannes, Alex... hope to see you soon, *come on!* ;)

Gracias a Ángel Carracedo, por tu confianza y apoyo en esta línea de investigación y por tu calidad humana excepcional. También muchas gracias a todos los compañeros de despacho y de trabajo durante estos años: Ricardo, Montse Camiña, Raquel, Montse Fernández y Xabi. Aquí Hugo obviamente se merece una nota muy especial: además permitirme trabajar en lo que más me gusta en unas condiciones fenomenales en tu laboratorio, has sido un excepcional consejero y compañero a lo largo de estos años. Siempre estaré agradecido por todo lo que he aprendido a tu lado. Ah, y un abrazo muy grande a Raquel, Mar y Lola! :)

Y como no podía ser de otra forma, muchísimas gracias a mi entorno familiar más cercano, empezando por mis padres José Manuel y Charo: me habéis dado muchas cosas, incluyendo la oportunidad de llegar hasta aquí, siempre he podido contar con vuestro apoyo incondicional. Igualmente, gracias a mis parientes a lo largo y ancho de España, algunos ya no están por desgracia, vuestro cariño ha sido importante. Mil gracias también a todos los miembros de la familia que gané al unirme con Tamara. Un especial recuerdo para Ángeles y Toño, para o Doutor Paulo, y para mi *ahijada política* Icíá ;)

Finalmente, si anteriormente me quedé corto en palabras de agradecimiento, ahora lo seré más. Tamara, sabes que eres y serás la persona más importante de mi vida, además de lo más maravilloso que me podría haber pasado. Sólo tú eres consciente de las fatigas (y momentos maravillosos, todo sea dicho) que hemos pasado juntos, la importancia de tu inmenso apoyo a lo largo de estos años es inconmensurable. Espero poder devolvértelo de alguna forma, de modo que al menos seas tan feliz como yo lo he sido gracias a ti.

ABSTRACT

G Protein-Coupled Receptors (GPCRs) constitute the most pharmacologically relevant superfamily of proteins. In this thesis, a computational pipeline for the structural modeling and the study of the dynamics of GPCRs is presented, and properly combined with experimental collaborations for the discovery and design of novel GPCR ligands. Our pipeline was first developed and applied for the characterization of the four subtypes of Adenosine Receptors, leading to the elucidation of ligand affinity and selectivity issues within this receptors family. Indeed, the employed implementation of Molecular Dynamics simulations has allowed the characterization of structural determinants of the activation process of adenosine receptors, and later in the assessment of the stability of GPCR dimers, in this case studying the CXCR4 receptor system. Both the activation and dimerization processes have great implications in the function and pharmacology of GPCRs, but we also provide specific applications in GPCR drug design projects. These include the discovery of novel scaffolds as potential antipsychotics, as well as the design of a new series of A₃ adenosine receptor antagonists, by using a successful combination of structure-based and ligand-based drug design approaches. Finally, the computational pipeline here developed has been integrated into a semi-automated computational protocol in the web server GPCR-ModSim (<http://gpcr.usc.es>), which is open to the scientific community. Altogether, the results of this thesis represent a relevant contribution to the structural biology and drug discovery of this superfamily of highly relevant pharmacological targets.

LIST OF PUBLICATIONS INCLUDED IN THIS THESIS:

Paper I: Gutiérrez-de-Terán H., Correia C., Rodríguez D., Carvalho M.A., Brea J., Cadavid M.I., Loza M.I., Proença M.F. and Areias F. (2009) Identification of novel scaffolds from an original chemical library as potential antipsychotics. *QSAR & Comb. Sci.* 28(8):856-60. <http://dx.doi.org/10.1002/qsar.200860198>

Paper II: Michino M., Abola E., GPCR Dock 2008 participants*, Brooks III C.L., Dixon J.S., Moutt J. and Stevens R.C. (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock. *Nat. Rev. Drug Discov.* 8:455-63. <http://dx.doi.org/10.1038/nrd2877>

* David Rodríguez is part of the consortium of participant researchers.

Paper III: Yaziji V., Rodríguez D., Gutiérrez-de-Terán H., Coelho A., Caamaño O., García-Mera X., Cadavid M.I., Brea J., Loza M.I. and Sotelo E. (2011) Pyrimidine Derivatives as Potent and Selective A₃ Adenosine Receptor Antagonists. *J. Med. Chem.* 54(2):457-71. <http://dx.doi.org/10.1021/jm100843z>

Paper IV: Rodríguez D., Piñeiro Á. and Gutiérrez-de-Terán H. (2011) Molecular Dynamics Simulations Reveal Insights into Key Structural Elements of Adenosine Receptors. *Biochemistry* 50(19):4194-208. <http://dx.doi.org/10.1021/bi200100t>

Paper V: Rodríguez D., Bello X. and Gutiérrez-de-Terán H. (2012) Molecular Modelling of G Protein-Coupled Receptors Through the Web. *Mol. Inf.*, 31:334-41. <http://dx.doi.org/10.1002/minf.201100162>

Paper VI: Rodríguez D. and Gutiérrez-de-Terán H. (2012) Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor. *Proteins* 80(8):1919-28. <http://dx.doi.org/10.1002/prot.24099>

The publications indicated above are reproduced in *Section 7* under the permission granted by the corresponding journals for this thesis (both in printed and electronic formats).

SUPPORTING INFORMATION:

Annex I: Rodríguez D. and Gutiérrez-de-Terán H. (2009) Adenosine Receptors: a systematic study of ligand binding based on the crystal structure of hA_{2A} receptor. Poster presentation at the *International Workshop In Memoriam of Ángel Ortiz*. Madrid, January 26-28.

Annex II: Rodríguez D., Sotelo E., Bender A. and Gutiérrez-de-Terán H.: Discovery of potent and selective adenosine receptor ligands via multi-objective design. Abstract of the oral presentation given at the *243rd ACS National Meeting & Exposition*. San Diego (California), March 25-29.

Annex III: SUMMARY IN SPANISH / RESUMEN EN CASTELLANO.

“Part of the inhumanity of the computer is that, once it is competently programmed and working smoothly, it is completely honest.”

Isaac Asimov

TABLE OF CONTENTS

1. INTRODUCTION	1
1.1. Biochemistry and pharmacology of G Protein-Coupled Receptors.....	2
1.1.1. Receptors involved in schizophrenia	3
1.1.2. Adenosine receptors	4
1.1.3. Chemokine receptors.	6
1.2. Structural Biology of GPCRs.....	7
1.2.1. Conserved sequence patterns and structural microswitches of GPCRs.....	8
1.2.2. Extracellular architecture of GPCRs.....	10
1.2.3. Advances in the structural characterization of GPCR oligomerization	13
2. MOLECULAR MODELLING METHODS.....	15
2.1. Computer-derived models of GPCR structures	15
2.1.1. GPCR modelling approaches	15
2.1.2. Online resources for GPCR modelling	17
2.2. Computational techniques of ligand discovery and design.....	19
2.2.1. Structure-Based approaches: ligand docking.....	19
2.2.2. Ligand-Based methods: linking chemical structure and bioactivity.....	20
2.3. Molecular Dynamics simulations	24
2.3.1. Force fields.....	24
2.3.2. Molecular Dynamics.....	24
2.3.3. Considerations of MD simulations of GPCRs	25
3. OBJECTIVES	29
4. RESULTS AND DISCUSSION.....	31
5. CONCLUSIONS.....	43
6. REFERENCES	45
7. PUBLICATIONS AND SUPPORTING INFORMATION.....	59

ABBREVIATIONS

5-HT: 5-hydroxytryptamine, serotonin
ADR: adrenergic receptor
AIDS: acquired immunodeficiency syndrome
APC: antipsychotic compound
AR: adenosine receptor
cAMP: cyclic adenosine monophosphate
CCR5: chemokine CC receptor 5
CV: cross-validation
CXCR4: chemokine CXC receptor 4
EL: extracellular loop
EPS: extrapyramidal side-effect
FDA: Food and Drug Administration (USA)
GDP: guanosine diphosphate
GPCR: G Protein-Coupled Receptor
GTP: guanosine triphosphate
HIV: human immunodeficiency virus
IL: intracellular loop
LB: ligand-based; LBVS: ligand-based virtual screening
LOO: leave-one out
LV: latent variable
MD: molecular dynamics
MIF: molecular interaction field
mGlu: metabotropic Glutamate receptor
MSeqA: multiple sequence alignment
MStA: multiple structure alignment
OR: opioid receptor
PC: principal component; PCA: principal component analysis
PDB: Protein Data Bank
PBC: periodic boundary conditions
PLS: partial least squares
PME: Particle Mesh Ewald
POPC: palmitoylolelphosphatidylcholine
QSAR: quantitative structure-activity relationships
RMSD: root-mean-square deviation
SAR: structure-activity relationships
SB: structure-based; SBVS: structure-based virtual screening
T4L: bacteriophage T4 lysozyme
TM: transmembrane helix
VS: virtual screening

1. INTRODUCTION

Currently, the pharmaceutical industry is facing a challenging scenario in the research and development of new drugs. The increasing costs and the difficulties in obtaining novel molecular entities demand an optimal performance in the different stages necessary for the eventual approval of a drug (1) (see Fig. 1).

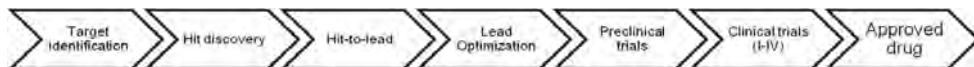


Figure 1. Schematic pipeline of the drug discovery and development process. The computational studies presented in this thesis will be focused in stages from hit discovery to lead optimization.

Computational methods are broadly integrated in the pipeline of drug discovery and development (2). Their application ranges several areas, including the management of databases and the prediction of physicochemical and pharmacokinetic properties, such as those related with absorption, distribution, metabolism and excretion (ADME) of chemical compounds in the body. Probably, the areas where computational techniques have demonstrated a broader impact are the discovery of novel hits, i.e. Virtual Screening (VS), and the rational design in lead optimization stages—e.g., using structure-based methods such as docking and molecular dynamics simulations—.

Here, some of these computational methodologies are applied in a deep study of the superfamily of G Protein-Coupled Receptors (GPCRs), a group of membrane proteins with high pharmacological interest. The application of a wide range of computational tools and protocols result not only in a deeper structural and dynamic characterization of these receptors, but also in the discovery and design of novel compounds targeting some of their members.

1.1. Biochemistry and pharmacology of G Protein-Coupled Receptors

G Protein-Coupled Receptors constitute the main superfamily of integral membrane proteins (3), the function of which is to transduce signals from the extracellular medium towards the cytoplasm. The canonical signalling mechanism of GPCRs involves the activation of the cognate heterotrimeric G Protein, located in the intracellular side of the receptor. Hence, the α subunit of this G protein exchanges a bound GDP with GTP, and dissociates from $\beta\gamma$ subunits upon their interaction with a (typically activated) GPCR. Different types of G proteins exist, where the characterized α subunits stimulate certain biochemical signalling pathways such as those involving cAMP and phosphatidylinositol (4), as illustrated in Fig. 2. Through this signalling diversity, GPCRs mediate a huge variety of physiological functions, which in part explains why they constitute the main drug target nowadays, accounting for at least 30% of the marketed drugs (5).

GPCRs are highly flexible receptors, exploring different conformations of varying functional significance, ranging from inactive (R_i) to active (R_a) forms of the receptor. The equilibrium between R_i and R_a in the apo form of the receptor dictates its basal activity. Extracellular ligands can influence this equilibrium in order to induce different biological responses, leading to an increase (full and partial agonists), maintenance (antagonists) or reduction (inverse agonist) of the basal activity of the receptor.

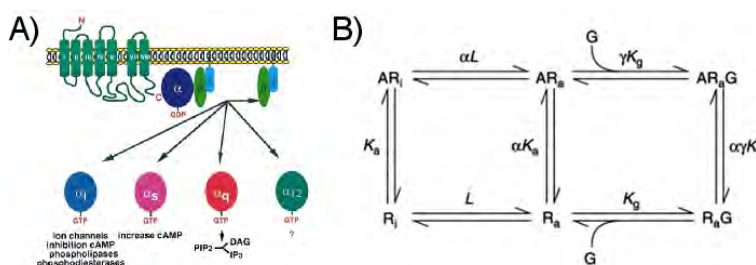


Figure 2. Signal transduction and conformational equilibrium of G Protein-Coupled receptors. **(A)** Cellular location of GPCRs and their interaction with cytoplasmic G Proteins, including examples of induced intracellular signalling cascades. Extracted from (6). **(B)** Extended Ternary Complex (ETC) model for GPCR activation, as originally proposed in (7). The different conformational equilibria for receptor activation, from R_i (inactive) to R_a (active), are represented. The explicit role of the binding of the natural agonist (A) and the G Protein (G) association is indicated as well.

Most of the GPCR families are characterized to bind a specific ligand, the natural agonist of the given receptor (7). Typically, GPCR ligands bind to the orthosteric site, the same of the natural agonist, located in the extracellular region of the receptor. Importantly, the higher sequence diversity among GPCRs is located in the extracellular half of the receptors (8). This is not surprising given the extremely diverse chemical nature of GPCR ligands: they include amines, nucleotides, peptides, proteins, lipids, organic odorants and ions. This is in line with the significant variety of physiological functions mediated by these receptors. Moreover, a classification of the GPCR superfamily, built on the basis of the pseudosequence formed by the hotspot residues within the binding crevice (9, 10) mostly agrees with the general phylogenetic tree of GPCRs (11), reflecting a balance in the evolution of ligands and receptors for the regulation of physiological processes.

A phylogenetic analysis of human GPCRs identified 5 groups of receptors, namely Glutamate, Rhodopsin, Adhesion, Frizzled/taste2, and Secretin clans, as defined in the so-called GRAFS classification scheme (11). Among these, the Class A/rhodopsin-like is the most populated group (700 out of the 800 characterized GPCRs). Class A receptors are further divided into four branches: α , β , γ and δ , which can be classified according to the similarity of the natural ligand (e.g., aminergic

receptors refer to all families which are activated by the endogenous biogenic amines), ultimately defining each family of receptors. The accumulated knowledge in medicinal chemistry on Class A GPCRs exemplifies their relevance among the whole superfamily. Indeed, the receptors considered in this thesis are members of rhodopsin-like GPCRs, and are introduced in the following subsections.

1.1.1. Receptors involved in schizophrenia

Schizophrenia is a psychiatric disorder that significantly affects the life quality of their patients, an estimated 1% of the worldwide population (12). It is a complex disease that, despite the constant advances in its pharmacological and biochemical characterization, still presents a challenging scenario for the discovery or design of novel compounds for its treatment.

The first antipsychotic compounds (APCs), termed as “typical”, were discovered during the middle of the 20th century (like chlorpromazine, see Fig. 3), and presented several therapeutic limitations including severe extrapyramidal side-effects (EPS) (13). Their antagonism on dopaminergic signaling was retrospectively discovered, and thus dopamine receptors were assigned as the first targets for the treatment of schizophrenia (14). However, latter works not only confirmed that observation, but also suggested the involvement of serotonin (5-hydroxytryptamine, 5-HT) receptors in the mechanism of action of novel antipsychotics (15). In fact, the relationship of affinities between serotonin and dopamine receptor subtypes was able to discriminate compounds with increased efficacy and less EPS: compounds showing a pK_i ratio between 5-HT_{2A} and D₂ receptors (*Meltzer Index*) above 1.2 presented this enhanced pharmacological profile, and were labeled as “atypical” APCs (15).

Despite the big efforts in the development of novel APCs, clozapine (see Fig. 3) remains as the gold standard in this regard after almost 50 years of its discovery (16). This drug presents the pharmacological advantages of atypical APCs, including reduced suicidal rate of patients (17); although it also involves several adverse effects such as agranulocytosis, seizures and weight gain (18). Several potential APCs have been introduced into clinical development stages in the recent years, which would mostly be considered as “me-too” congeners of clozapine (19).

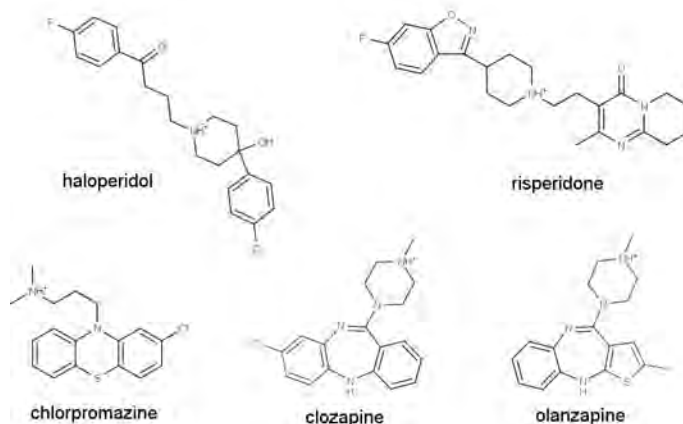


Figure 3. 2D structures of representative APCs. The positively-charged nitrogens at physiological pH are indicated accordingly.

The paradigm for the mechanism of action of APCs has evolved throughout the years into the current “magic shotgun” model of Roth and co-workers (20): compounds are hoped to present a specific pharmacological profile (including activity, efficacy and even pharmacokinetics considerations) against a panel of targets. This contrasts with the much earlier “magic bullet” analogy formulated by Ehrlich, where only the action of a compound against a single target is considered.

Thus, a portion of the *receptorome* is necessary to understand the mode of action of APCs, involving many subtypes of families of aminergic receptors (dopamine, serotonin, histamine, muscarinic, α and β adrenergic...) and other GPCRs such as metabotropic Glutamate (mGlu) receptors. The ideal APC should be able to achieve the appropriate output in dopaminergic signaling in specific regions of the brain, by means of achieving a specific efficacy balance in a battery of receptors, while avoiding or minimizing the affinity in those targets related to side-effects. However, the modelling of such process is extremely complex. Novel APCs departing from known scaffolds towards new chemistries, at the same time complying with the desired pharmacological profile, are hoped to provide better tools for the future treatment of this elusive disease.

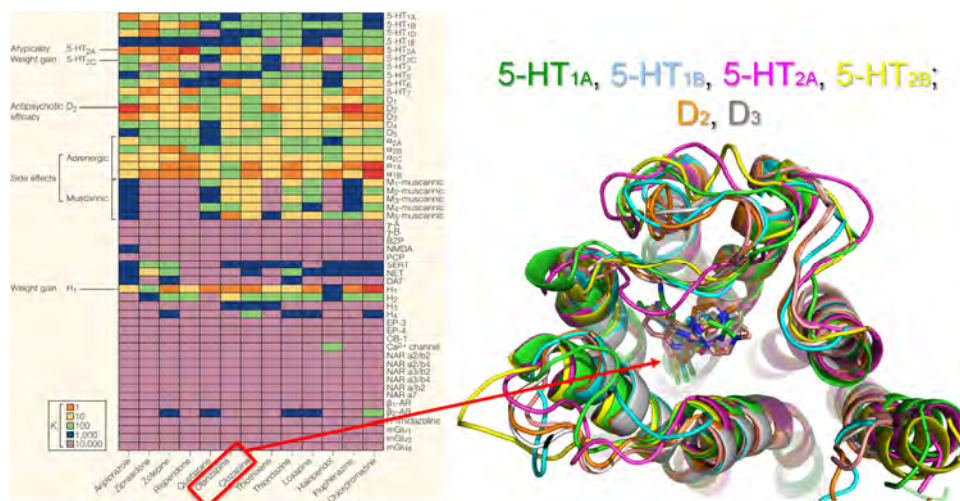


Figure 4. Receptors involved in schizophrenia. On the left hand side, the different affinity constants (K_i) of several APCs (columns) against a panel of targets involved in this disease (*receptorome*, in rows) are indicated, together with associated efficacies and side-effects of the receptors —chart taken from (20)—. On the right hand side, homology models and X-Ray structures of several serotonin and dopamine receptor subtypes are shown, together with computationally-predicted binding modes of the APCs clozapine and olanzapine. This is part of an ongoing study of our laboratory.

1.1.2. Adenosine receptors

The ubiquitous nucleoside adenosine regulates a wide range of biological functions through the activation of a family of GPCRs, the so-called adenosine receptors (ARs), which are subdivided in four subtypes: A_1 , A_{2A} , A_{2B} and A_3 (21). ARs mediate the typically cytoprotective function of adenosine, leading to a different response to its variable extracellular levels depending on the corresponding tissue and physiopathological state (22). The four subtypes present different couplings with intracellular effectors, and consequently their activation transduces diverse intracellular signaling pathways. On one hand, A_1 and A_3 subtypes (sharing a 49% sequence identity) couple to G_i proteins, inhibiting the activity of adenylyl cyclase upon activation. On the other hand, A_{2A} and A_{2B} (59% seq. id.) interact with G_s proteins, eventually leading to the reduction of intracellular cAMP levels (22). At the same time, despite their relatively ubiquitous body distribution, the presence of A_1 and A_{2A} (high affinity for adenosine) is mainly focused in the nervous system and cardiovascular tissues, meanwhile A_{2B} and A_3 (low affinity for adenosine) are generally found in peripheral tissues such as kidney or lungs.

ARs are highly relevant drug targets due to their extensive potential in a wide variety of pathologies. Their modulation has interesting indications, and drug design efforts have been

traditionally focused in the discovery of orthosteric ligands chemically similar respect to natural products that bind ARs (see below). However, the development of potent and selective drugs within this family is quite challenging due to the structural similarity of human ARs. Moreover, interspecies differences of these receptors have hampered the extrapolation of results from preclinical studies in animal models to humans. In fact, apart from the use of adenosine in the treatment of tachycardia, only regadenoson is a FDA-approved drug targeting an AR (see Fig. 5). Furthermore, this A_{2A} AR selective agonist is used as a diagnostic tool in myocardial perfusion imaging, rather than as a therapeutic agent (23).

Regarding the efficacy of ligands targeting adenosine receptors, ARs agonists are specially interesting for the treatment of cardiovascular diseases such as arrhythmia or vasodilation processes (22), as well as for contributing to the preconditioning of cardiomyocytes in ischemia through the activation of A_1 and A_3 subtypes (24). The structure-activity relationships (SAR) around the adenosine scaffold have been thoroughly explored (25), exemplified by the selective A_{2A} AR agonists CGS21680, UK-432079 and the aforementioned regadenoson (see Fig. 5).

Following the same rationale, the design of novel ARs antagonists was initially focused on substitutions on the xanthine scaffold (26), illustrated by the weak and non-specific natural products caffeine and theophylline (see Fig. 5). Many ARs antagonists are applied in nervous system disorders, where caffeine exerts stimulant and enhanced awareness effects. In fact, an inverse relationship between coffee intake and the development of Parkinson's disease has been observed (27), and selective A_{2A} AR antagonists such as preladenant or tozadenant (with a non-xanthine scaffold) are undergoing clinical trials for the treatment of this disease (28). ARs antagonists are also used in pathologies of peripheral tissues, including the development of selective A_{2B} AR compounds as anti-asthmatic preclinical candidates (29) or A_3 AR antagonists for the treatment of glaucoma and osteoarthritis already in initial clinical trials (30).

As exemplified above, intensive efforts are being carried out in the drug development of ARs chemical modulators. A special effort is given to the design and synthesis of compounds with selective profiles and novel chemical scaffolds towards this therapeutically underexploited family of receptors.

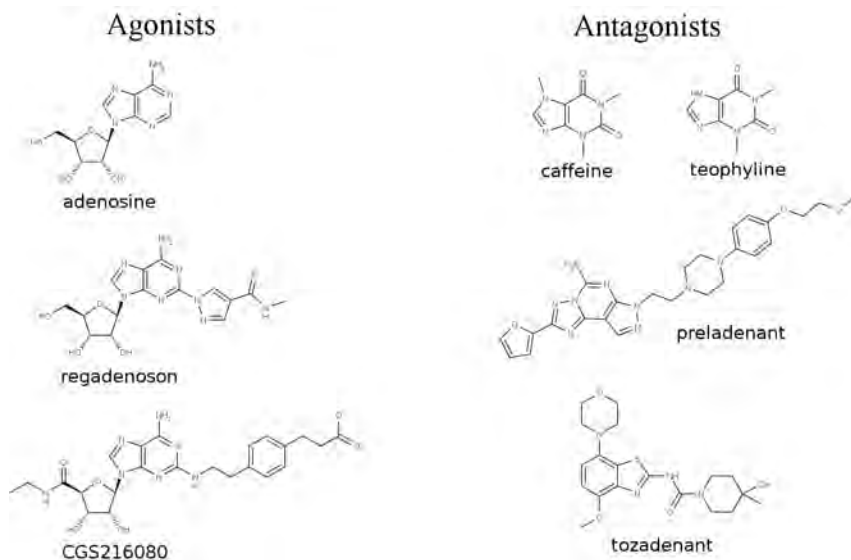


Figure 5. Representative ARs ligands, grouped according to their efficacy.

1.1.3. Chemokine receptors.

Chemokines are small proteins that induce the chemotaxis of somatic cells upon the activation of chemokine receptors, which eventually promote the necessary dynamics of the cytoskeleton. Up to 20 members of chemokine receptors, a family of GPCRs, have been described for humans. They are typically implicated in cell migration processes and related pathologies such as inflammation and autoimmune diseases (31). One of the most studied member of this family is CXCR4, which is activated by the chemokine CXCL12 (31) and participates in the cell entry process of HIV-1 in the host cell (32-34). Additionally, CXCR4 has been involved in the development of more than 20 types of cancer (35), thus offering a wide range of potential therapeutic applications.

Drug design efforts on CXCR4 have been focused in the development of antagonists for blocking the HIV-1 entry (36). The virus internalization process starts with a complex formed by gp120 and CD4, glycoproteins of the virus and host cell respectively (37). Two chemokine receptors are described to act as co-receptors, forming an eventual heterotrimeric complex necessary for the membrane fusion process: *i)* CCR5 is primarily used by M-tropic HIV-1 strains in the asymptomatic phase of the disease, and is targeted by the antagonist maraviroc, an FDA-approved drug for the treatment of AIDS (38). *ii)* CXCR4 is mainly employed by T-tropic strains of the virus, predominant in the latter and rapid progress stage of the disease (39), providing a crucial complementation for drug resistance and receptor switching mechanisms of CCR5-based therapies (36). Initial efforts in the design of CXCR4 inhibitors were focused on screening peptides resembling the natural chemokines, eventually leading to clinical candidates (40, 41). Lately, small organic compounds have been developed, including successful design efforts starting from a bicyclam scaffold exemplified by plerixafor (AMD3100) (42), giving rise to the Phase I compound AMD070 (43) (see Fig. 6). However, the lack of highly potent chemical scaffolds towards these receptors is still a challenging issue, due to the difficulty of disrupting the protein-protein interactions (gp120 / CXCR4) with small organic ligands (44).

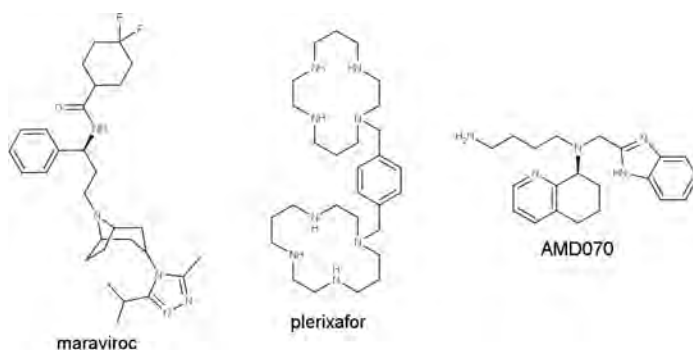


Figure 6. 2D structures of representative compounds targeting chemokine receptors.

1.2. Structural Biology of GPCRs

On one hand, the obtaining of high-resolution structures of GPCRs is an extremely difficult task due to the inherent flexibility of these receptors, and their lack of stability when they are extracted from their lipidic environment. On the other hand, the topology of GPCRs is one of their characteristic and more conserved features: a helical bundle constituted by 7 transmembrane helices (TMs), connected by 3 extracellular (EL) and 3 intracellular (IL) loops. Generally, an additional helix (H8) runs perpendicular to this bundle in the C-terminus region. This topology has been recognized for decades and has significantly contributed to the initial generation of GPCR structural models using bacteriorhodopsin as a template (45), followed by the trace of C α atoms of frog rhodopsin obtained by electron cryomicroscopy 3D maps (46). Finally, the crystallographic structure of bovine rhodopsin was released in the year 2000 (47), being the unique template of atomic resolution for GPCR modelling during the subsequent seven years (see Section 2.1). Thereafter, the development of paradigmatic technologies for receptor stabilization and crystallization (48, 49) opened a new era in the structural biology of GPCRs. Two of the most successful approaches consider the fusion of bacteriophage T4 lysozyme (T4L) in the IL3 (50, 51), or the introduction of residue point mutations in order to increase the thermal stability of the receptors (49, 52). So far representative members of two branches, five sub-branches and eight GPCR families of rhodopsin-like receptors have been crystallized in their inactive conformation (see Table 1). In addition, members of three of these families (rhodopsin, adenosine and β -adrenergic receptors) have also been solved in an active-like conformation.

		Active < -----> Inactive		
		Ligand [PDB code] (<i>reference</i>)		
Species	Receptor	Agonist	Antagonist	Inverse agonist
Human	A _{2A} AR	NECA [2YDV] (53) Adenosine [2YDO] (53) UK-432097 [3QAK] (54)	T4G [3UZA] (55) T4E [3UZC] (55)	ZM241385 [3EML] (56) Caffeine [3RFM] (57) XAC [3REY] (57)
Turkey	β 1ADR	Carmoterol [2Y02] (58) Isoprenaline [2Y03] (58) Dobutamine [2Y00] (58) Salbutamol [2Y04] (58)	Cyanopindolol [2VT4] (52) Iodocyanopindolol [2YCZ] (59)	Carazolol [2YCW] (59) Bucindolol [4AMI] (60) Carvedilol [4AMJ] (60)
Human	β 2ADR	BI-167107 [3SN6] (61) FAUC50 [3PDS] (62)	Alprenolol [3NYA] (63)	Carazolol [2RH1] (50) Timolol [3D4S] (64) JSZ [3NY9] (63) ICI118551 [3NY8] (63)
Human	CXCR4		IT1t [3ODU] (65) CVX15 [3OE0] (65)	
Human	D ₃ R		Eticlopride [3PBL] (66)	
Human	H ₁ R		Doxepin [3RZE] (67)	
Human	M ₂ R		QNB [3UON] (68)	
Mouse	M ₃ R		Tiotropium [4DAJ] (69)	
Mouse	δ -OR		Naltrindole [4EJ4] (70)	
Mouse	μ -OR		β -FNA [4DKL] (71)	
Human	κ -OR		JDtic [4DJH] (72)	
Human	NOPR		C-24 [4EA3] (73)	
Human	S1P ₁ R		ML056 [3V2Y] (74)	
Bovin	Rhodopsin	Trans-retinal [3PQR] (75) Apo-Opsin [3CAP] (76)		Cis-retinal [1U19] (77) β -ionone [3OAX] (78)
Squid	Rhodopsin	Trans-retinal [3AYM] (79)		Cis-retinal [2Z73] (80)

Table 1. Summary of the GPCR crystal structures available to the scientific community. Receptor-ligand structures are categorized on the basis of the biological response induced by the co-crystallized ligand. Only the structure with the highest resolution (typically below 3 Å) is indicated for those ligand-receptor complexes crystallized more than once.

Finally, the first GPCR—G Protein complex has been solved for $\beta 2$ adrenergic receptor ($\beta 2$ ADR) with its cognate G_s protein (61). The coordinates of the available GPCR structures can be found in the Protein Data Bank (PDB, <http://www.pdb.org>), the reference database in structural biology (81). The most representative GPCR-ligand complexes solved to date (with their corresponding PDB codes) are compiled in Table 1, more details can be found in recent reviews of the field (82, 83).

Despite the described experimental advances, the structural coverage of GPCRs is still sparse (see Fig. 7). Two out of four branches of Class A receptors remain uncharacterized (β and γ), despite the pharmaceutical interest of some of their members. The β branch includes targets such as neuropeptide Y (84) and gonadotropin-releasing hormone receptors (85), apart from others currently employed in clinical studies such as oxytocin hormone receptor (86). The populated family of purinergic P2Y receptors, attractive targets for diseases such as platelet aggregation (87), comprise one of the most relevant members of γ branch. Regarding α and δ branches, their structural coverage is below the 10% of their members (excluding olfactory receptors). In this scenario, computationally derived models of good quality are highly demanded to deepen the structural knowledge of the immense majority of Class A receptors, as well as for the rest of GPCR classes.

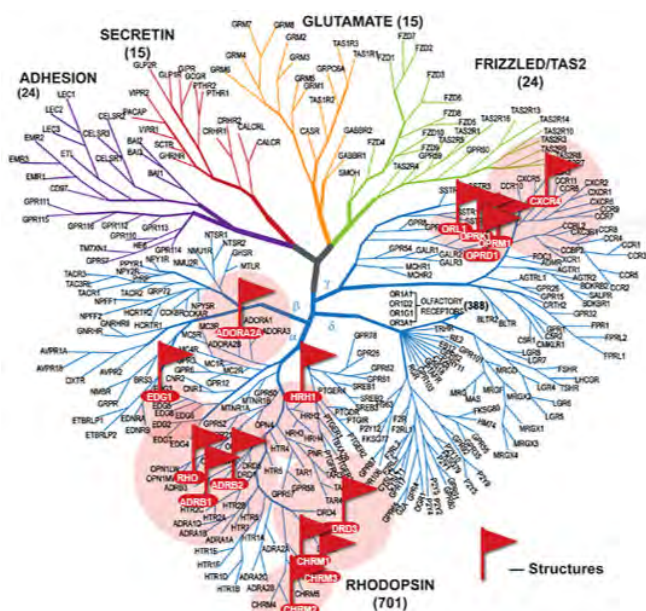


Figure 7. Structural characterization of GPCRs in the context of their phylogenetic tree. Each red flag marks a receptor that has been solved by X-Ray crystallography. This figure is extracted from the NIH GPCR Network website (<http://gpcr.scripps.edu/>, accessed on September 2012).

1.2.1. Conserved sequence patterns and structural microswitches of GPCRs

Several patterns have been traditionally observed in GPCR sequence alignments, exemplified by the Ballesteros & Weinstein scheme for residue numbering (88). Following this proposal, residues are numbered with the X.YY code, where X represents the TM helix, and YY is a correlative number according to the sequence of the receptor, assigning the number 50 to the most conserved GPCR residue in the corresponding helix. This scheme is adopted along this thesis (in uppercase) and related works. The first hallmark of GPCRs is the high conservation of proline residues in different TMs, including TM5, TM6 and TM7. These residues produce distortions in the canonical α -helix

secondary structure that induce kinks and hinges, found to be important for the structure, dynamics and function of membrane proteins (89).

The most conserved sequence patterns of GPCRs involve functionally relevant regions, ultimately characterized as key structural elements of GPCR crystallographic structures. The general location of the most prominent functional microswitches can be found in Fig. 8, and are described below:

- The *ionic lock*. This *lock* is formed by salt bridge established between Arg^{3.50} [part of the conserved D(E)RY motif] and a D/E^{6.30} residue, at the intracellular tips of TM3 and TM6 respectively. It has been generally considered that inactive conformations of GPCRs might present this *ionic lock* formed, as already proposed in rhodopsin structures and spectroscopic measurements (90, 91). The disruption of such interaction should be necessary to expose the appropriate intracellular surface for interacting with the cognate G Protein upon activation (92).
- The rotameric *toggle switch* involves the side chain conformational change of the conserved Trp^{6.48}. The side chain rotamer transition from *gauche+* to *trans* has been largely hypothesized from rhodopsin experiments (93), site-directed mutagenesis, sequence conservation analysis (94) and computational calculations (95) regarding its implication in the activation pathway of GPCRs. Trp^{6.48}, located in the bottom tip of the orthosteric GPCR binding site, is part of the conserved CWxP motif in TM6. It is supposed to modulate the corresponding helix kink induced by Pro^{6.50}, "straightening" the conformation of TM6 and contributing to the separation of its intracellular tip from TM3 (see the *ionic lock* point above) (92, 95, 96).
- The NPxxY motif at TM7, includes the conserved Pro^{7.50} that induces a marked kink in the intracellular part of the corresponding helix. Also, Asn^{7.49} is part of a *GPCR conserved hydrogen bond network* involving residues from TM1-TM2-TM6-TM7, including Asn^{1.50}, Asp^{2.50}, and Asn^{7.45}. The interactions between the aforementioned residues are supported by a tight network of structural water molecules revealed by crystallographic GPCR structures, which are hypothesized to be involved in the regulation of the *toggle switch* and other aspects of receptor activation (97-100). Recently, the last crystal structure of the A_{2A}AR has confirmed that this site also hosts a sodium ion, which acts as an allosteric modulator in many GPCRs (101).

Initial hypothesis regarding the involvement of the *ionic lock* and the *toggle switch* in the activation process of GPCR have been under debate, especially after the release of recent X-Ray structures (83). The generally broken *ionic lock* in inactive-like structures has been intriguing (48, 82), although it is partially in agreement with the observed conformational rearrangements in active-like structures: a considerable separation of intracellular tips of TM6 respect TM3 occurs upon activation, as evidenced in the GPCR-G_s protein crystallographic complex (61). Thus, the relevance of the closed *ionic lock* in inactive-like conformations of GPCRs remains unsolved only considering static crystallographic structures. Respect the *toggle switch*, the fact that the active-like crystal structures do not display major conformational changes of Trp^{6.48} suggested new concerted mechanisms for activation (102, 103), while recent inactive-like crystal structures showing a rotamer more closer to *trans* (68, 69, 71, 72) might support the hypothesis about a transient conformational change along the activation pathway (94, 104).

The study of crystallographic conformations of the aforementioned microswitches (with special attention to the *ionic lock*), together with the stability of the conserved GPCR hydrogen bond network, have been studied and rationalized by means of MD simulations. The reader is referred to Section 4 and Paper IV for details in this regard.

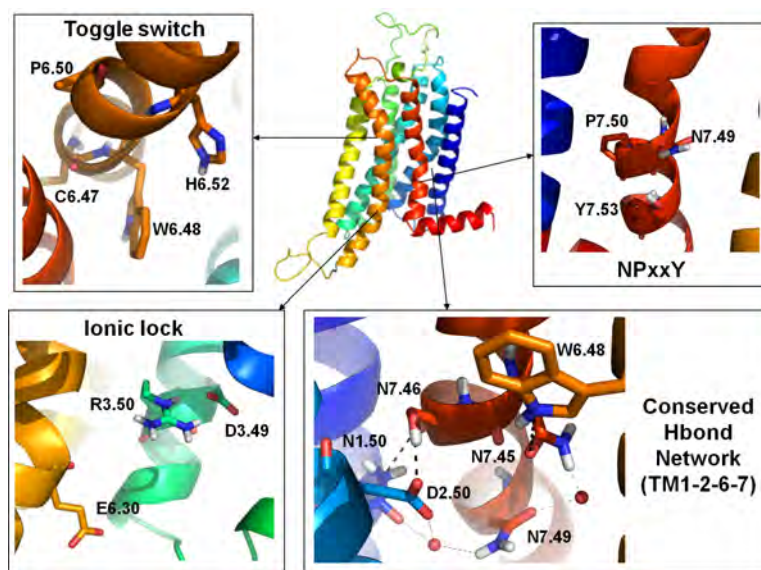


Figure 8. Structural microswitches of GPCRs. These sites are exemplified by the structure of A_{2A} AR, with its residues coloured from blue (Nt) to red (Ct). The location of the toggle switch, the ionic lock, the NPxxY motif and the conserved network of residues at TM1-TM2-TM6-TM7 are depicted.

1.2.2. Extracellular architecture of GPCRs

The functional microswitches presented in the previous section are located in the intracellular half of GPCRs (except the *toggle switch*, in between both halves), as they are involved in conformational changes necessary for the activation of the receptor, and the eventual signal transduction through cytoplasmic effectors. On the other hand, the extracellular region of GPCRs, where the ligand binding occurs, presents a high sequence and topological diversity (82). All crystallographic structures, except rhodopsin, show open conformations of the 3 ELs (see Fig. 9). This is because rhodopsin function is mediated by a covalently-bound ligand (retinal), buried in the binding site, which is isomerized by photons in order to activate the receptor. Rhodopsin receptors have been widely employed for the theoretical elucidation of GPCR structures and activation mechanisms (105). However, the rest of GPCRs have to bind bioorganic molecules that stabilize the active state, and present significantly different extracellular topologies in order to accommodate the binding of the corresponding ligands, typically in a buried and hydrophobic site. In all cases, the EL2 (between TM4 and TM5) shows the most variable conformations among all GPCR crystallographic structures, as it presents the longest and most variable sequence length among the ELs. In fact, peptide-binding receptors present an exposed region favoured by a hairpin-shaped EL2 (with a beta-sheet segment), in order to adapt bulkier ligands compared to small organic compounds (see Fig. 9C). This can be observed in crystallographic structures of CXCR4 (65) and the four opioid receptors (ORs) solved to date (70-73). A very relevant anchoring point delimitates two sections of the EL2, made by the conserved disulfide bridge in Class A receptors. This bridge is formed between a cysteine located in EL2, and the other (Cys^{3.25}) in TM3. The section between this disulfide bridge and TM5 generally delimitates one of the upper tips of the binding sites of GPCRs, and even establishes specific contacts with the co-crystallized ligands for adenosine, β 2ADR and D₃ receptors (see Table 1). Additional disulfides are observed between several ELs, where the paradigmatic case of A_{2A} AR shows disulfides between EL1-EL2, and within EL2 and EL3 respectively (56), further constraining the conformational space of these flexible regions.

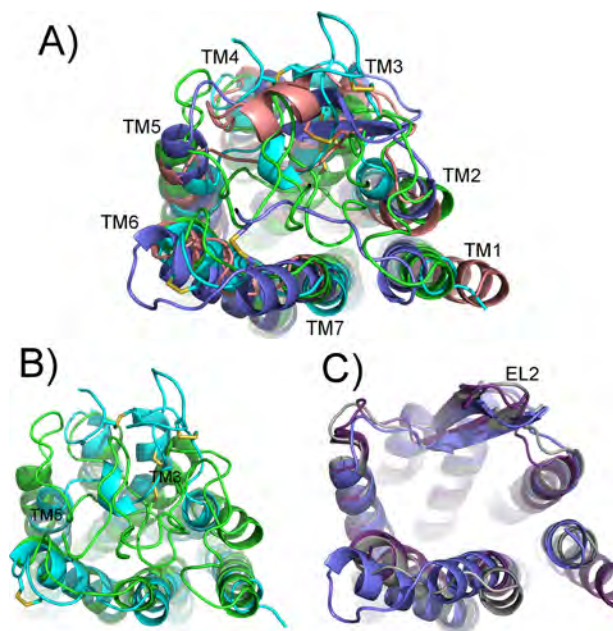


Fig. 9. Comparison of the extracellular topology of crystallographic GPCRs. **(A)** The general arrangement of the ELs is shown for rhodopsin (green), $A_{2A}AR$ (cyan), $\beta 2ADR$ (brown) and CXCR4 (dark blue). The location of TMs is labelled, and disulfide bridges between adjacent cysteines are shown in sticks. **(B)** Comparison between rhodopsin and the small molecule-binding receptor $A_{2A}AR$. One of the most remarkable differences is located in the section of EL2 between TM3 (where the GPCR-conserved disulfide bridge is located) and TM5. **(C)** The particular conformation of EL2 of peptide-binding receptors, with an extended β -sheet, is exemplified with the structures of CXCR4, κ -OR (violet) and μ -OR (grey).

The extracellular binding site of GPCRs is a deep hydrophobic cavity, thus considered as “druggable” (106) and generally presenting excellent opportunities for computational drug design (107). The ligands typically bind to a subsite involving TM3-TM5-TM6-TM7, as observed in the crystallographic structures and already inferred by mutagenesis and sequence analyses (9). In order to provide context of the studies presented in this thesis, we will briefly describe the main structural considerations regarding GPCR ligand binding in the light of the different crystallographic structures of $\beta 2ADR$, $A_{2A}AR$ and CXCR4 receptors (see Fig. 10).

The first human GPCR to be crystallized was $\beta 2ADR$, in complex with the inverse agonist carazolol (50). Extensive hydrophobic interactions of the ligand in the binding site are complemented with three polar contacts, especially the salt bridge established with the conserved Asp^{3.32}, together with hydrogen bonds with Ser^{5.42} and Asn^{7.39}. The abovementioned salt bridge has been traditionally known to be formed with a positively charged nitrogen of aminergic receptor ligands (see Fig. 10A). This structure supposed a significant boost for the structural characterization of GPCRs at the time of its release (see below). Latter structures of biogenic amine receptors in complex with antagonists presented different sets of receptor-ligand interactions around the conserved salt bridge (59, 66-73). Alternatively, structures of β -adrenergic receptors have been crystallized with inverse agonists, antagonists and agonists (see Table 1), showing subtle differences in rotamers of serines at TM5, already anticipated by computational methods (108).

In 2008, the crystallographic structure of $A_{2A}AR$ in complex with the potent antagonist ZM241385 was solved (56). The main ligand-receptor interactions involve a double hydrogen bond with Asn^{6.55} and a π -stacking interaction with Phe^{5.29}, both residues being totally conserved in ARs family. Additionally, polar contacts are established with Glu^{5.30} (see Fig. 10B). This structure offers an

extremely valuable picture of the main hallmarks of antagonist ligand binding on ARs, further explored by computational methods (109, 110), including Papers III, IV and Annex I. Later on, structures of A_{2A}AR in complex with xanthine derivatives (57), novel antagonists (55) and synthetic (54) and natural (53) agonists were solved with different crystallization methods (see Fig. 10D for the case of UK-432097 agonist). The key interaction with Asn^{6.55} has been observed in all cases, and the involvement of Ser^{7.42} and His^{7.43} in agonist binding (through hydrogen bonds with the ribose moiety of the ligands) was confirmed. All these data is in agreement with previous mutagenesis studies (26, 111, 112).

Recently, the structures of CXCR4 in complex with two antagonists, the cyclic peptide CVX15 and the small organic ligand IT1t, were solved (65). Importantly, the aforementioned structures show the formation of symmetric CXCR4 homodimers (see next section for details). As discussed above, the ligand binding site presents a large surface and open disposition in a relatively solvent-exposed region (65). CVX15 show extensive polar contacts with the receptor, highlighting the interactions of arginines 1, 2 and 14 of the ligand with residues Asp^{5.26} (at EL2), Thr^{3.33}, Asp^{4.60}, Asp^{6.58} and His^{3.29} of the receptor. On the other hand, the nitrogens of the isothiourea group of IT1t show two potential salt bridges with residues Asp^{2.63} and Glu^{7.39} (see Fig. 10C). Both the peptidic and the organic ligand occupy an overlapping binding site, which is markedly different to the one previously observed for rhodopsin, β₂ADR and A_{2A}AR receptors, illustrating the structural plasticity of GPCRs for ligand binding (65).

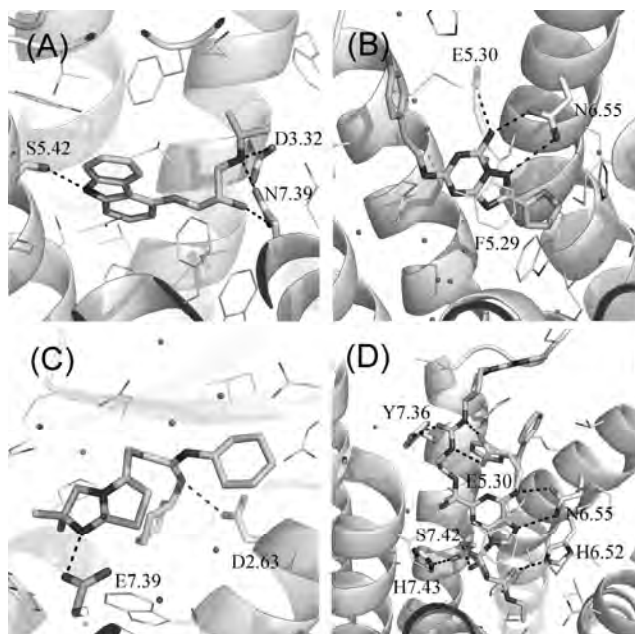


Figure 10. Comparison of the binding sites of relevant crystallographic GPCR complexes. The selected complexes are (A) β₂ADR – carazolol, (B) A_{2A}AR – ZM241385, (C) CXCR4 – IT1t and (D) A_{2A}AR – UK-432097 (see Table 1 for details). Polar contacts between receptors and ligands are indicated with dashed lines, and the involved residues are labelled.

1.2.3. Advances in the structural characterization of GPCR oligomerization

A phenomenon of increasing interest in the field of GPCR research is the growing evidence of dimerization and oligomerization processes in their function. Early studies shown that dimers of receptor chimeras were able to recover both ligand binding and function upon their interaction (113). Since then, the study of the biologically relevant quaternary structure of GPCRs has been pursued by a plethora of biochemical and pharmacological studies (114). The possible suprastructural architectures could involve several constituents (from dimers to oligomers of varying number), being either formed by the same (homo-) or different protomers (heteromers). These phenomenon offers new possibilities for the understanding of the signal diversification mediated by GPCRs, with consequent pharmacological applications (115). Paradigmatic examples of interacting GPCRs involve homodimers with implications in nervous system disorders such as 5-HT_{2A}-mGluR in schizophrenia (116), or A_{2A}AR-D₂ in Parkinson's disease (117).

Initial studies on the structural determinants of GPCR oligomerization relied on bioinformatics approaches (118), together with the semi-empirical model of rhodopsin oligomers (119, 120). This model constituted the most accepted GPCR dimerization mode for several years, involving the interaction of TM4 and TM5 between protomers (121); as well as contacts between TM1, TM2 and TM7 in order to complete oligomerization, observed in later rhod-opsin X-Ray structures (76, 122). Afterwards, the high resolution crystal structures of CXCR4 with different antagonists (see above) supposed a breakthrough in the field as they presented a novel GPCR dimerization interface. Interestingly, homodimerization of CXCR4 has been demonstrated as a key phenomenon in the biological function of this receptor (123). In fact, its ligand-independent homodimerization has been characterized by fluorescence and bioluminescence resonance energy transfer techniques (124, 125). Moreover, models of the possible stoichiometries of CXCR4-CXCL12 complexes (natural agonist binding), and gp120-CD4-CXCR4 heterotrimers (HIV-1 internalization process) have been suggested in the light of the here discussed crystallographic homodimers (65). All five solved structures of CXCR4 (PDB codes 3ODU, 3OE6, 3OE8, 3OE9 and 3OE0) show a similar dimerization interface (65). Protomers are arranged in a parallel and symmetric disposition, compatible with the orientation in the cellular membrane, and presenting a significant buried surface area ($\approx 850 \text{ \AA}^2$). All these indications supported the potential biological relevance of the observed dimerization mode (118), where the main hydrophobic interactions involved TM5 (residues Leu^{5.33}, Val^{5.36}, Val^{5.37}, Phe^{5.40} and Met^{5.44}), with additional symmetric hydrogen bonds between residues in the extracellular tips of TM5 and TM6 of both protomers (Asn^{5.31}-Leu^{6.62}; Asn^{5.31}-Glu^{6.64}; Trp^{5.34}-Leu^{6.63}). Very recently, the crystal dimers of κ -opioid receptor (72) show similar contacts respect the TM1-TM2-TM7 interface observed in rhodopsin structures (see above). More importantly, an analogous dimerization mode compared to the one observed in CXCR4 structures has been observed in the crystal of μ -opioid receptor (71). In this case, the solved homodimer shows a higher buried surface area (up to 1300 \AA^2) between protomers, with an interface involving more extensive contacts between TM5 and TM6 (see Fig. 11).

The conjunction of biophysical experiments (126) with more reliable structural models of GPCR dimerization and oligomerization is hoped to increase the accuracy of the molecular understanding of these protein-protein interactions. Interesting applications include the design of bivalent and dimer-specific ligands, in order to design more selective drugs (121).

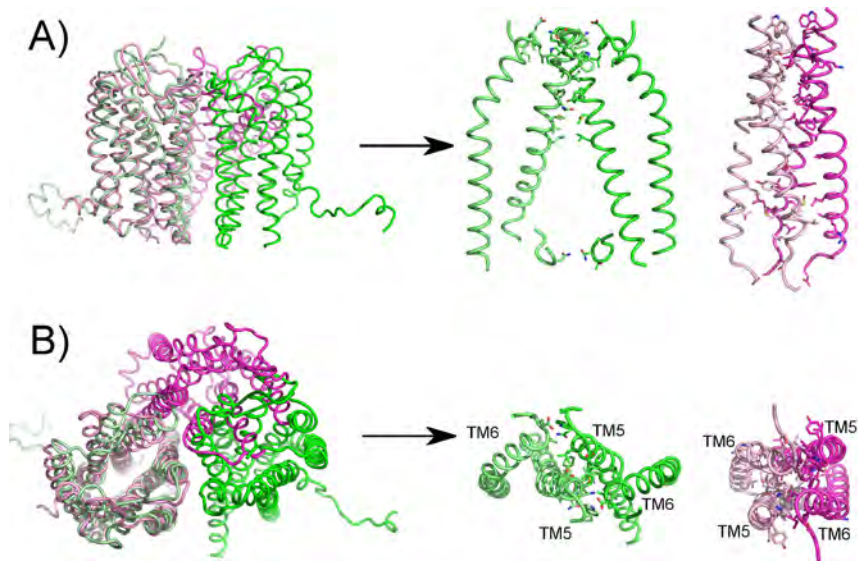


Figure 11. Comparison of the dimerization mode of CXCR4 (green) and μ -OR (magenta) crystallographic structures. In the right hand side, general disposition of the dimers are shown in lateral (**A**) and top (**B**) views. For each point of view, specific interactions involving TM5 and TM6 (and IL2 for CXCR4) are highlighted. Residues establishing contacts with the opposite protomer are shown in sticks.

2. MOLECULAR MODELLING METHODS

In this section, the main computational techniques for protein structure prediction, ligand design and molecular dynamics are outlined. A special stress is made on the methods employed through this thesis, together with particular considerations of their application in the field of GPCRs.

2.1. Computer-derived models of GPCR structures

2.1.1. GPCR modelling approaches

Despite the high sequence diversity among GPCRs, reliable computational 3D models of these receptors can be obtained taking advantage of the conserved 7TM topology in the whole superfamily, with applicability in different GPCR structure-based drug design projects (83). Three main techniques are distinguished for the prediction of the 3D structure of GPCRs (127):

- Topology-based techniques (also known as *ab initio* modelling) have been specifically developed for the case of GPCRs (128, 129). The different TMs are built and lately packed reconstructing the conserved topology of a 7TM bundle using force field-based biophysical calculations, e.g., MD simulations. Historically, this was the first methodology to produce GPCR models on the basis of bacteriorhodopsin (45) and early rhodopsin structural information (130). Several programs with different protocols have been developed and employed throughout the years (129, 131, 132), being MembStruck the most extensively used to date (128, 133).
- Threading (fold recognition) techniques evaluate the suitability of different templates and map the sequences of the target and templates, thereafter a range of assembly refinement methods are generally applied in order to build the final model (134).
- Homology modelling: this procedure builds the 3D structure of the target protein on the basis of a sequence alignment with a template of known structure. Spatial restraints, derived from the query-sequence alignment and the 3D structure of the template, guide the initial modelling of the target, which is followed by additional refinement stages (135, 136).

The different modelling protocols have been evaluated in recent Critical Assessment of GPCR Structure Modelling and Docking (GPCR Dock) competitions (137, 138). Here, in a similar fashion as CASP and CAPRI contests, researchers were asked to submit computer-generated models of a query GPCR-ligand complex, prior to the release of the corresponding crystal structure. *GPCR Dock 2008* was the first call, employing the A_{2A}AR in complex with the potent antagonist ZM241385 (56) as test case. Our laboratory took part in this competition (see *Section 4* and specifically Paper II). A latter contest was performed in 2010 (138), where the structures of D₂ (66) and CXCR4 (65) receptors in complex with diverse antagonists were selected for the challenge. The different modelling procedures were evaluated for their ability to reproduce the structure of the target receptor, as well as the prediction of the native contacts with the co-crystallized ligand by means of ligand docking methods (see *Section 2.2.1*). Comparative modelling techniques emerged as the most popular ones, and indeed showed the best performance.

As it could be deduced from the GPCR Dock contests and the vast literature on the field, the application of several circumventions in GPCR homology modelling are typically adopted (127, 139).

Specifically, *i*) a special attention is given to conserved motifs in TM helices, as they govern the most relevant part of the template-query alignment (46, 88). *ii*) The modelling of the loops should be considered as a separate issue, in particular the long EL2 often involved in ligand binding (140). *iii*) Some refinement of the model is advised, such as ligand-biased models (127, 141, 142) or other biophysical methods for incorporating flexibility of the receptor (143) including MD simulations (144-146). The selection of the best template and the use of one or several combined templates has also been assessed (139, 147-149), although it has been observed that the consideration of multiple templates would just slightly improve the use of one unique structure (147).

Homology modelling of GPCRs has been traditionally limited because of the reduced availability of appropriate templates, taking into account that a minimum threshold of 30% sequence identity between query and template is generally accepted (150). However, the new crystallographic information (82) and the conserved 7TM topology offers excellent starting points for producing high quality homology-derived models, specially for their application in the design of orthosteric ligands (83, 137, 138). Moreover, we estimate that currently 58% of the pharmacologically relevant human GPCRs (excluding olfactory and orphan receptors) present the minimum 30% of sequence identity in the TM region with at least one of the available templates (see Fig. 12). The inherent limitations of homology modelling methods, mostly due to the relative structural bias introduced by the template(s), has motivated the development of first principle methods (see above). Still, the performance of these methods did not show a clear superiority respect to homology-based methods (132, 137, 138). Encouragingly, homology-derived structures have shown their utility in SBVS (151, 152) and/or ligand design (152) projects. Also, the usefulness of such models in retrospective virtual screenings has been demonstrated when only a few crystal examples were available (127), and in a very recent analysis on β 2ADR homology models (153).

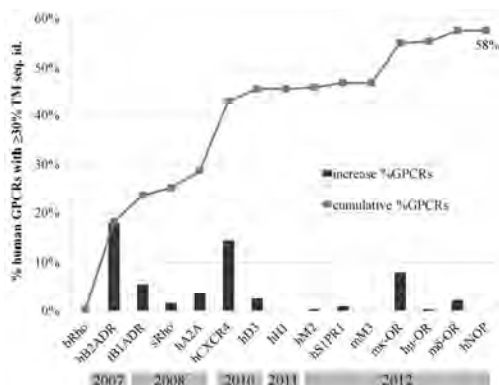


Figure 12. The impact of crystal structures in the homology modeling of human GPCRs. A threshold of 30% sequence identity in the TM region is accepted to define a valid template for a given receptor. Orphan and olfactory receptors have been excluded from this analysis, in order to reflect the impact on the receptors with highest pharmacological interest. For each new template, the vertical bars represent the percentage of new human GPCRs that could be reliably modelled only after the release of that template (which date is indicated in horizontal bars). This analysis has been performed with the GPCR-ModSim web server (see below).

In this thesis, homology modelling has been the selected technique for predicting the structure of the GPCRs under study: the serotonin receptor 5-HT_{2A} (Paper I) and all the four members of the AR family (Annex I, II and Papers II—IV). A specific protocol has been designed in these studies, being finally implemented in an automated fashion within a web-server developed and maintained by our laboratory (see Paper V). This protocol consists on the following steps:

- Sequence alignment with the considered template. This is performed with the software ClustalX2 (154) using PAM250 substitution matrices. The revision of the sequence alignments in the context of receptor families and manual refinements are performed when necessary, with special attention to the correct alignment of TM helices and disulfide bridges.
- Generation of homology models with the software MODELLER (155), using the aforementioned target-template sequence alignment. Additional optimization of the conformation of loops with the loopmodel routine (156) was considered in Papers **III**, **IV**.
- The selection of the output models is carried out attending to the energetic DOPE-HR scoring implemented in MODELLER, and to stereochemical quality reports produced by PROCHECK (157) and Molprobit server (158).
- Molprobit is also employed for the addition of hydrogen atoms, optimizing their orientation based on potential hydrogen bond networks between residues. Additionally to Molprobit, PDB2PQR (159) and Schrödinger utilities (160) were employed for the assessment of tautomeric and protonation states of titratable residues. More exhaustive calculations of this kind were carried out with MCCE software (161) in Paper **IV**.
- An energy minimization of the model is performed in order to relax the structure, avoiding possible clashes, with molecular modelling software suites such as MOE and Schrödinger (160). These optimizations are performed by algorithms that evaluate the negative gradient of the potential energy of the system with force fields, see *Section 2.3.1* for details.
- Finally a validation of the models is assessed with molecular docking approaches (see *Section 2.2*). This includes the agreement with mutagenesis data in ligand binding (Papers **I**, **II**), as well as the selectivity profiles of reference compounds (performed in *Annex I* and further used in Papers **III**, **IV**).

2.1.2. Online resources for GPCR modelling

Reliable comparative modelling can be performed by the automation of the necessary steps: selection of the best template(s), query-target sequence alignment, and generation of the homology model. This is exemplified by servers and repositories available in the web, such as ModBase (162) or SWISS-MODEL (163). However, this general purpose pipelines for homology modelling do not capture the specific nuances of membrane protein modelling in general, and GPCRs in particular. Taking the special features of these receptors into account, several web servers and databases have been elaborated for the case of GPCRs. They are reviewed and compared with the GPCR-ModSim web server (<http://gpcr.usc.es>), a service developed in the context of this thesis, in Paper **VI**.

Attending to the modelling technique, two resources employing derivations of the threading methodology of TASSER are available. GPCR-ITASSER takes into account protein-membrane interactions and mutagenesis restraints compiled in the GPCR Restraint Database (GPCRRD) (164). Very recently, TASSER^{VMT}-lite has been presented as well (165), providing a database of generated models of all human GPCRs.

Web toolkits using homology modelling protocols are more popular, and some of them have been developed in the recent years. MEDELLER employs its own methodology, and considers the

environment of the lipid bilayer for the generation of coordinates of membrane proteins (166), although the user must provide the template structure, and its sequence alignment with the query receptor. On the other hand, GPCR-SSFE (148) employs MODELLER, and incorporates the possibility of employing several templates for the generation of the TM bundle of the target receptor. Additionally, the GPCR DataBase (GPCRDB) (167) —a valuable resource for sequence, structure and mutagenesis data on these receptors— has recently deposited homology models generated with the software YASARA.

We have adapted the employed GPCR homology modelling protocol into the GPCR-ModSim web server. The capabilities of this service include the sequence alignment of the query receptor against a profile of the available templates, aiding the selection of the most appropriate template. Afterwards, the initial homology models can be built, followed by an additional loop optimization stage. Finally, the main novelty of this service is the possibility of performing MD simulations of the generated models. For a more details about the protocol, capabilities and performance of GPCR-ModSim, the reader is referred to *Section 4* and *Paper V*.

The adoption of more accurate methodologies for receptor modelling, together with their easier access to researchers without extensive experience in molecular modelling, is hoped to enhance the utility of computational models of GPCRs in the drug research and biochemical fields.

2.2. Computational techniques of ligand discovery and design

Here, a brief depiction of the computational approaches for ligand design employed in this thesis will be outlined, which can be mainly subdivided in structure-based (SB) and ligand-based (LB) methods. SB approaches take into account direct information of the structure of the receptor, meanwhile LB methods are explicitly unaware of it. Finally, proper combinations of SB and LB methods are especially suitable in drug design, an strategy that is carried out in Papers **I**, **III**, Annex **II**, and further outlined in following sections.

2.2.1. Structure-Based approaches: ligand docking

The spontaneous association of a ligand and a receptor in order to form a non-covalent complex is produced when the associated free energy of binding (ΔG_{bind}) is negative. The more negative the binding free energy, the tighter the ligand-receptor association is. This can be related to thermodynamic experimental observables such as the protein-ligand dissociation constant (K_i). Two main components of binding free energy are distinguished: the enthalpic (ΔH) and the entropic ($-T\Delta S$) terms, both containing solute and solvent contributions. Benefitting from the high-resolution structures of protein-ligand complexes (compiled in databases such as the PDB) and the available ligand affinity data, several methods have been developed in order to estimate the ligand binding affinities, given a reliable prediction of the correct geometry of the protein-ligand complexes. A widely employed approach in this regard is molecular docking (168), where docking software programs perform two main tasks: *i*) the generation of the possible conformations of the ligand in the binding site of the receptor and *ii*) the assessment of the latter with a scoring function. Different types of these mathematical functions are distinguished:

- **Force field-based:** the binding energy is calculated from non-covalent intermolecular interactions by means of force fields, see *Section 2.3.1*.
- **Empirical:** they consist on linear combinations of empirically-defined binding terms derived from experimental affinities.
- **Knowledge-based:** they employ statistical potentials trained with intermolecular interactions extracted from 3D structure databases.

Either explicitly or implicitly, non-covalent interactions between the ligand and the protein are taken into account, including salt bridges, hydrogen bonds, van der Waals and hydrophobic contacts. Scoring functions might also consider other components of the binding energy, such as desolvation or entropic terms (169, 170). The predicted affinities can be used to computationally rank a chemical library according to the docking scores (structure-based virtual screening, SBVS) in order to prioritize the experimental ligand-binding assays; or the obtained docking poses can be employed for rationalizing chemical substitutions of compounds in lead optimization efforts (2).

Molecular docking considering full flexibility of the ligand, and a rigid receptor (except the rotamers of polar hydrogens) has been employed in several works of this thesis: for the GPCR Dock 2008 contest (see above and Paper **II**), and in order to provide a rationale of the binding affinity and selectivity of novel ligands on the basis of the predicted binding modes (see Annex **I** and Papers **I**, **III**). From the available docking software (169), GOLD (171) was employed in the different studies, together with the empirically-derived ChemScore (172) scoring function. ChemScore estimates the total free energy change upon ligand binding as:

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}} \quad \text{Eq. (1)}$$

Each component of the equation is calculated on the basis of different physical considerations of binding. An empirical scaling factor, extracted from a regression on binding and crystallographic

data, is included in each of them. The final ChemScore is obtained by adding a clash penalty and internal torsion terms, in order to avoid close contacts and conformations of bad stereochemical quality. An optional covalent term can be added, not considered in the calculations here presented.

$$\text{ChemScore} = \Delta G_{\text{binding}} + P_{\text{clash}} + c_{\text{internal}}P_{\text{internal}} + (c_{\text{covalent}}P_{\text{covalent}} + P_{\text{constraint}}) \quad \text{Eq. (2)}$$

GOLD employs a Genetic Algorithm (GA) for the exploration of conformations of the ligand in the receptor. A population of initial random solutions is encoded as chromosomes, containing information regarding the physical complementarity of atoms of the ligand and the protein, in terms of establishing potential interactions (hydrogen bonds donors and acceptors, hydrophobic points...). Afterwards, a series of iterative optimizations are carried out, in a process emulating the evolution of populations. The fitness of the “chromosomes” is evaluated by the scoring function in every stage, until the generation of the final solutions and their corresponding scores, according to the selected parameters.

The exponential growth of available GPCR structures (82) has been translated into successful prospective SBVS by means of high-throughput docking on several crystallized GPCRs of different families (107). Novel pharmacological hits have been discovered with highly-enriched computational screenings of large chemical libraries on β 2ADR (173, 174), A_{2A}AR (110, 175), D₃ (151), CXCR4 (176) and H₁ (177) receptors. Interestingly, homology models were simultaneously employed in two of these studies, with similar performance and with complementary solutions compared to the corresponding crystallographic structures (151, 176). Thus, the increased structural knowledge of GPCRs, together with appropriate computational modelling of receptors of unknown structure, is hoped to provide further lead compounds with pharmacological applications in several GPCR families.

2.2.2. Ligand-Based methods: linking chemical structure and bioactivity.

The traditional lack of structural information within the GPCR superfamily has motivated the extended use of ligand-based (LB) drug design techniques to assess the medicinal chemistry of GPCRs. Indeed, despite the recent increase in the number of GPCR crystal structures, the accumulated experience on LB methods still provides a valuable tool in the process of ligand design (178). These methods rely on the molecular similarity principle, extensively employed in computational drug discovery and design, following the assumption that similar compounds might share comparable physicochemical and/or pharmacological properties, such as bioactivity (179). The final goal of these approaches is to come up with novel chemical entities with desired characteristics, usually by building and rationalizing structure-activity relationships (SAR). The general procedure for measuring molecular similarity by computational methods starts with *i)* the numerical representation of the chemical structure with *descriptors*, followed by *ii)* a *metric measure* that evaluates the eventual chemical likeness. Several methods in this regard have been developed throughout the years (179, 180). In this thesis we will focus on three specific methodologies: shape-based virtual screening, 3D quantitative structure-activity relationships (3D-QSAR) and molecular similarity searching on the basis of atom environments.

2.2.2.1. Shape-based virtual screening

Shape-based VS is a 3D-based computational method, the aim of which is to retrieve molecules with similar spatial features compared to a known *query* molecule (sort of encoding key binding characteristics) but hopefully with different structure, thus opening *scaffold hopping* opportunities by means of bioisosteric replacement (181). In this section, the basic concepts of shape-based VS are introduced in the context of the protocol followed in Paper I with software from OpenEye suite (www.eyesopen.com).

First, it is necessary to generate energetically-accessible 3D-conformers for each ligand, including the chemical database to be screened and the query compound. In our case, we employed

the software OMEGA, which is well known for its ability to reproduce bioactive-like conformations of organic molecules (182). Afterwards, the Rapid Overlay of Chemical Structures (ROCS) program is used for VS by superimposing all conformers from the chemical database against the query—which in fact can be made of several molecules or even receptor grids—. The method employs a Gaussian-based function that produces molecular alignments maximizing the overlap between the compared molecules with a remarkable computational performance (183). The molecular alignment can be refined with a so-called color force field (not to confuse with Molecular Mechanics force fields described in Section 2.3.1), which evaluates the relative disposition of chemical groups with similar functionalities (hydrogen bond donors and acceptors, hydrophobic and charged groups, and rings). Finally, a combined score (ComboScore) is calculated for every screened compound, on the basis of the shape and chemical similarity with respect to the reference molecule, following the formula:

$$\text{ComboScore} = \text{Tanimoto}_{A,B} + \text{ColorScore}_{A,B} \quad \text{Eq. (3)}$$

The $\text{Tanimoto}_{A,B}$ and $\text{ColorScore}_{A,B}$ respectively measure the shape and chemical overlay of the reference (A) and screened (B) compound. Both take values from 0 to 1, and thus the ComboScore ranges from 0 to 2. On one hand, the shape superimposition is calculated with a Tanimoto coefficient:

$$\text{Tanimoto}_{A,B} = \frac{O_{A,B}}{I_A + I_B - O_{A,B}} \quad \text{Eq. (4)}$$

The I terms are the self overlaps of each molecule, and $O_{A,B}$ represents the overlap between A and B. On the other hand, the ColorScore measures the superposition of chemically similar groups (see above) using a hard Gaussian function in a sphere around each atom.

The final computational step involves the visual inspection of the top-ranked molecules, according to their ComboScore, with the software VIDA (184). A special attention is paid to the overlays with the reference molecule, and finally the most promising compounds are selected for pharmacological evaluation.

2.2.2.2. 3D-QSAR

Unlike shape-based virtual screening, Quantitative Structure-Activity Relationships (QSAR) methods are focused on elucidating the structural determinants of bioactivity in series of relatively close analogs (185). Again, several descriptions and representations of the molecules (mainly 2D or 3D) can be considered. In this thesis, a 3D-QSAR approach has been employed in Paper III, using Molecular Interaction Fields (MIFs) as descriptors. This methodology follows the work of Goodford for characterizing protein binding sites with the GRID methodology (186). The aim of this approach is to obtain differences in the calculated fields between active and inactive molecules. Thus, a regularly-spaced 3D grid is defined in the surroundings of each compound. Then the energy of interaction between a chemical probe and the molecule is calculated in each point (node) of the grid with a Molecular Mechanics function. The chemical probes represent chemical groups of different nature, i.e., hydrogen bond donor and acceptor groups (N1 and O probes, respectively), hydrophobicity (DRY) and shape (TIP). These probes can be selected for the representation of potential interactions of the ligand with the target receptor. Finally, the calculated energies in each node of the grid are encoded into arrays, and the combination of the calculated arrays for each molecule forms a matrix which is ready to be analyzed with multivariate statistical analyses, as outlined below.

A key step for building classical QSAR models is the molecular alignment of the compounds in space. One option is to employ a biological superimposition of the ligands in the target receptor, generally by means of molecular docking (187). In parallel, several descriptors have been developed in order to represent the most relevant regions for ligand-receptor interactions, such as GRid Independent Descriptors (GRIND), including their latest implementation GRIND-2 (188). Although the

structural alignment of the molecules is not strictly necessary for the GRIND-2 methodology, it has been shown to increase the accuracy of the method. Thus, both the *docking-based* alignment of the ligands and the GRIND-2 descriptors were our reference methodology in Paper III.

Given the extensive number of variables generated by MIF descriptors for each molecule, it is necessary to apply statistical methods in order to obtain correlations between structural variations of the ligands and their biological activities. **Principal Component Analysis** (PCA) is a multivariate analysis method that aims the reduction of the dimensionality of data, with the hope of discovering trends in sets of objects (in this case molecules) described by variables (from descriptors) given an initial data matrix. The initial variables are replaced by fewer new variables, so-called Principal Components (PC). These PCs must be orthogonal, i.e., they cannot be correlated, and thus the first PCs retain most of the information (variance) of the initial data. **Partial Least Squares** (PLS) is a regression method related to PCA, widely employed in 3D-QSAR. It is used to correlate sets of variables—in this case the ones generated by MIF descriptors—with the biological activity of the compounds. By linear combination of the initial variables, few Latent Variables (LV) are obtained in order to explain variations in activity. Taken the experimental values, a correlation coefficient (r^2) with the model can be calculated. Additionally, cross-validation (CV) is typically performed for discarding possible over-fitting issues originated from spurious selection of variables. In Paper III, Leave-One-Out (LOO) was the selected CV method, where each compound is extracted from the set and predicted by a model formed by the rest of objects. Several metrics can be employed in this regard, where the predictive correlation coefficient (q^2) is widely used:

$$q^2 = 1 - \frac{(y - y')^2}{(y - \bar{y})^2} \quad \text{Eq. (5)}$$

Where y and \bar{y} are the experimental (bioactivity) values and their average respectively, and y' are the predicted values by the model. Typically, one should evaluate these statistics in models with different numbers of LVs, selecting the one where r^2 and q^2 start to reach a plateau. This practice ensures a good compromise between the descriptive ability of the model and precluding the possible over-fitting of variables against bioactivity.

2.2.2.3. Molecular similarity with atom environments

This LBVS technique relies on the 2D description of molecules in order to find new compounds with similar properties respect to compounds with desired characteristics, following the molecular similarity principle described above. The particularities of the method MOLPRINT2D (189, 190), used in this thesis in a Virtual Screening on adenosine receptors (see Section 4 and Annex II) are here presented. As other similarity searching approaches, MOLPRINT2D can be summarized in three steps:

- **Definition of the molecule.** In this case, MOLPRINT2D employs atom environments of the molecules as descriptors, derived from the connectivity table. Sybyl MOL2 atom types are assigned to heavy atoms, describing their physicochemical properties. The corresponding fingerprints are calculated for every heavy atom, encoding their environment with all other atoms lying within to 2 bonds.
- **Feature selection.** Given two subsets of molecules of different classes, i.e., active and inactive compounds, the information gain measure of Quinlan (191) is employed for obtaining the most relevant features for their separation.
- **Classification.** A Naïve Bayes classifier (see Eq. 6) is finally employed for the scoring of new molecules, assigning a probability of class membership (active or inactive) according to the features and classes of the training sets (representing a vector connecting the chemical space from inactive to active molecules).

$$\frac{P(CL_1|F)}{P(CL_2|F)} = \frac{P(CL_1)}{P(CL_2)} \prod_i \frac{P(f_i|CL_1)}{P(f_i|CL_2)} \quad \text{Eq. (6)}$$

Where two classes of molecules, 1 and 2 (e.g., active and inactive) and i features (from the feature selection step) are considered. F is a vector formed by the feature elements f_i , and $P(CL_1)$ and $P(CL_2)$ are respectively the frequencies of active and inactive molecules. Consequently, the conditional probabilities $P(f_i|CL_n)$ represent the relative frequencies of the features in each active and inactive classes. Finally, the probability of a new molecule to belong to either class is obtained by the first ratio of Eq. 6, calculated according to its corresponding feature vector F .

MOLPRINT2D has proven an impressive performance in the retrieval of active molecules compared to contemporary molecular similarity searching approaches (189, 190, 192). The increasing availability of bioactivity data of chemical compounds from public repositories (193) is hoped to provide enhanced opportunities for building bioactivity models with the potential of discovering new pharmacological hits.

2.3. Molecular Dynamics simulations

2.3.1. Force fields

MD simulations are based on a molecular mechanics description of the system. In this framework, a force field function defines different interactions of the particles (atoms) in order to assess the potential energy of the system:

$$\begin{aligned}
 U_{pot} = & \sum_{bonds} \frac{k_b}{2} (b - b_0)^2 + \sum_{angles} \frac{k_\theta}{2} (\theta - \theta_0)^2 + \\
 & + \sum_{dihedrals} k_\phi [1 + \cos(n\phi - \delta)] + \sum_{impropers} \frac{k_\xi}{2} (\xi - \xi_0)^2 + \\
 & + \sum_{non-bonded} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}^2} + \sum_{non-bonded(LJ)} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)
 \end{aligned} \tag{Eq. 7}$$

The main interactions between atoms can be subdivided into bonded and non-bonded. The first type includes bond stretching, angles, dihedrals and impropers between atoms linked from 1 to 3 bonds. Regarding non-bonded interactions, these are divided into electrostatic and non-electrostatic, respectively modelled by Coulomb and Lennard-Jones potentials. Given the general force field definition (see Eq. 7), the variables (distances and angles) represent relative spatial positions of the atoms in a given configuration of system. The indicated constants (*parameters*) are derived in order to reproduce observables from exhaustive computational calculations (e.g. quantum mechanics), or experimental values such as spectroscopic experiments or free energies of solvation. Each force field has its own formulation and parameterization scheme, as well as different compatibilities with the available MD software packages (194). The OPLS force field, developed by the group of Jorgensen (195), has been employed throughout this thesis.

2.3.2. Molecular Dynamics

The main aim of molecular dynamics (MD) simulations is to reproduce the dynamic evolution of the structure of biomolecules in order to gain insights into their function. Most of the high-resolution structures deposited in the PDB have been obtained by X-Ray crystallography (196). This technique has the limitation of providing frozen snapshots —indeed, the temperature is generally lowered for favouring protein crystallization— of systems that otherwise are in solution (or within a cellular membrane, as is the case of GPCRs) and under the effect of thermal motion at physiological temperature. In order to sample the thermally-accessible configurations of a biomolecular system, MD simulations rely on force fields and statistical mechanics in order to reproduce macroscopic properties while observing microscopic details. This way, several explorations can be made with this technique, including the assessment of the stability of protein structures, the characterization of conformational rearrangements, or the study of protein-ligand binding determinants.

Molecular dynamics follow the rules of Newtonian motion, where the potential energy of the atoms is calculated with the force field function. Each atom of the system is given an initial coordinate (provided by the structure) and velocity (assigned following a Maxwell-Boltzmann distribution). Then, forces can be derived from the negative gradient of potential energy.

$$F_i = -\nabla_i V(\mathbf{r}) = -\frac{\partial V(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \tag{Eq. 8}$$

Where V is the force field-derived potential energy, and F_i is the force on the atom i , according to the positions (\mathbf{r}) of the atoms of the system (from i to N). Afterwards, the new coordinates of the atoms are calculated attending to the integration of Newton's second law of motion:

$$F_i = m_i a_i = m_i \frac{\partial^2 r_i}{\partial t^2} \quad \text{Eq. (9)}$$

The update of coordinates and velocities of the particles is typically performed by the leapfrog algorithm:

$$r_{(t+\partial t)} = r_{(t)} + v_{(t)}\partial t + a(\partial t)^2/2 \quad \text{Eq. (10)}$$

$$v_{(t+\partial t)} = v_{(t)} + \frac{(a_{(t)}+a_{(t+\partial t)})\partial t}{2} \quad \text{Eq. (11)}$$

The ∂t is discretized in timesteps, typically in the order of 1 or 2 femtoseconds, for the integration of the equations of motion. The calculations described so far are iteratively repeated for a given number of steps. Consequently, coordinates, velocities and energies of the system are regularly collected into the so-called MD trajectory, describing the evolution of a biomolecular system in time with atomistic detail.

Since the systems should be treated as infinite, but the computational representations of the modelled system must be necessarily finite, several models exist in order to avoid spurious effects from the boundaries of these simulated systems. Here, the periodical boundary conditions (PBC) are a common practice, and are employed in the works presented in this thesis. In PBC, the simulation box has a given geometry, and adjacent images "see" each other, thus representing a continuous and infinite space: if a molecule crosses one boundary of the box, it will enter from the opposite edge, thus conserving the total number of particles in the system. The PBC representation is associated with other approximations, in particular for the evaluation of non-bonded interactions corresponding to atoms which are separated more than a given distance cutoff, which cannot be estimated with the regular force field terms outlined in Eq. 8 due to an exponential increase in computational time. A common technique for the computation of long-range electrostatic interactions beyond such cutoff is the Particle Mesh Ewald (PME) summation method (197). Other approximations that allow a reasonable computation time is the use of bond constraints for the bonds of hydrogen atoms (in our case, the LINCS method). Finally, there are several formal considerations in order to obtain thermodynamic ensembles, such as the coupling to a temperature bath and to a barostat in order to obtain a canonical ensemble (NPT), used through the works here presented. These and other technical issues are detailed in the methods section and supplementary material of Papers IV, VI, as well as in the GROMACS manual (198), which was the software used for MD simulations throughout this thesis.

2.3.3. Considerations of MD simulations of GPCRs

Membrane proteins are given increasing interest in the biochemical and pharmacology fields, as long as they are in charge of numerous cellular communication functions. Not only receptors (such as GPCRs), but also transporters and channels are targets of pharmaceutical interest. Consequently, it is important to decipher the molecular determinants of their function by studying relevant conformational rearrangements, and their interaction with other molecules such as lipids or small organic compounds (199). Given that the function of GPCRs is governed by a dynamic conformational equilibrium between active and inactive states (7, 102, 200), MD simulations are an appropriate technique to explore, at least partially, such conformational equilibrium. Moreover, these simulations offer an excellent complementation of the increasing structural information provided by the recent crystallographic efforts on GPCRs, combined with additional biophysical techniques (201-203).

Thus, several computational methods for MD simulations of membrane proteins have been developed in the recent years (204, 205). They facilitate the preparation of the initial systems, which typically model the biological environment of the protein with a lipid bilayer and solvent molecules. Added to the increase in computational resources and methodologies, longer timescales can be reached (from hundreds to thousands of nanoseconds in all-atom simulations) in order to explore

relevant conformational rearrangements in the activation pathway of GPCRs, as exemplified by recent works in the field (206-208).

Within the framework of this thesis, a novel computational protocol for membrane protein insertion and equilibration has been developed and presented in detail in Paper IV, and further extended and automated in Papers V, VI. Several considerations of this procedure, which is based on the GROMACS suite of programs, are briefly explained and outlined in Fig. 13:

- A palmitoyloleoylphosphatidylcholine (POPC) bilayer pre-equilibrated at 260K,¹ thus in crystalline phase, in order to help the packing of the lipids around the protein.
- The GPCR of interest is inserted into the bilayer, and overlapping lipid molecules are removed. The protein must be correctly aligned in the Z-axis, where TM4 is supposed to run parallel to the lipid tails, and H8 should cross the intracellular polar heads, in agreement with the assigned orientations of GPCRs deposited in the OMP database (209). Manual refinement of the position in the Z-axis is performed in order to obtain a balanced number of lipids in the upper and lower leaflets of the bilayer.
- Finally, the system is solvated with water molecules and neutralized with counter-ions.

The geometry of the simulation box is a hexagonal prism, which optimizes the number of solvent molecules for saving computational time (with an estimated reduction that exceeds the 10% as compared with a regular rectangular-shaped box). The typical size of the final simulation box is constituted by around 60.000 atoms for protein monomers (Papers IV, V) and 100.000 atoms for protein dimers (Paper VI).

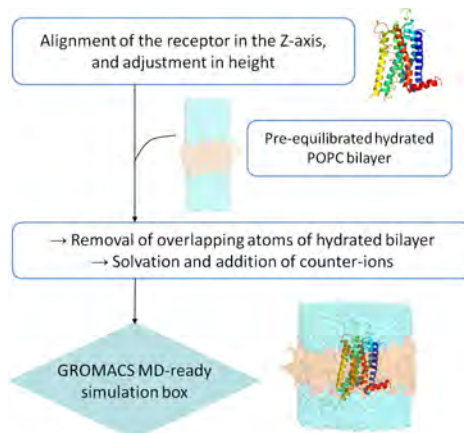


Figure 13. Outline of the computational protocol for the setup of MD simulation boxes of GPCRs. The different stages for the insertion of the GPCR structure, here exemplified by an A_{2A}AR monomer, are depicted. More details can be found in the main text and Papers IV—VI.

The MD simulations force fields (see above) are parameterized for proteins, but not for other molecules such as organic ligands or lipids. However, it is necessary to obtain proper parameters for all the members of the system for a given force field to be employed in the simulations. For the case of small organic compounds, we generated reliable parameters for the selected OPLS-AA force field (195) with the molecular modelling software Schrödinger (160), a procedure that proved to be valid in

¹ POPC bilayers have been employed in Papers IV—VI. All simulations were run at 310K, where POPC transitions towards the biologically-relevant fluid phase.

recent protein-ligand binding free-energy calculations (210). For lipid molecules, one alternative is their reparameterization (211) from *ad-hoc* lipid force fields, such the one developed by Berger *et al.* (212). Another option is to adapt two force fields (one describing the lipid, and the other the protein/ligands), as proposed by the half- ϵ double pairlist method (213) developed for the GROMACS simulation package. This method ensures a proper scaling of 1-4 non-bonded interactions (those involving atoms separated by 3 consecutive bonds) for both the Berger (lipid) and OPLS-AA (protein) force fields, while successfully reproducing experimental macroscopic properties such as the area per lipid (214). We have used this methodology in Papers **IV**, **VI**, as well as in the MD pipeline of GPCR-ModSim server (Paper **VI**).

Once the system is set up, an equilibration scheme is necessary in order to allow the different constituents to adapt to each other, as the achievement of a stable configuration is necessary to obtain meaningful equilibrium simulations. In equilibration stages, the positions of certain protein atoms (and ligands and crystallographic waters where considered) are restricted by a force constant, which applies an energetic penalty for the movement of the restrained atoms. These constants are gradually reduced until the whole system is unrestrained. Then the production phase of the simulations starts, where the collection of data takes place. Several equilibration protocols have been explored, taking especial care of the loop regions in Paper **IV**. In that same work, independent replicas of selected systems were simulated in order to increase the thermodynamic sampling, and to gain statistical significance on the observed dynamic events.

The monitoring and analysis of MD trajectories is a key and time-consuming step. With the purpose of facilitating and automating this task, we developed a set of scripts that take care of the trajectory visualization (using the program PyMOL, www.pymol.org) and the calculation of geometrical and energetic observables, mainly using several tools in the GROMACS package as well as plotting programs (GNUplot, XMGrace). Regarding the latter calculations, they are important to assess the stability of the system and to quantify conformational changes and interactions of interest. Some of these analyses have been adapted for the special case of GPCRs, and implemented into the GPCR-ModSim server (see *Section 4* and Paper **V**).

3. OBJECTIVES

The main motivation of this thesis is to provide a methodology for characterizing the structure and dynamics of selected members of GPCRs, aimed to the elucidation of the mechanism of receptor activation, as well as in drug discovery and design efforts. We have pursued these challenging goals through the establishment of three milestones, which are here outlined and further analysed on the basis of our results in the next section:

- *Development and application of a robust homology modelling methodology in GPCR computer-aided ligand discovery and design.*

This milestone would be accomplished through the consideration of the following objectives: *i)* development of reliable structural models of selected GPCRs of interest, through a methodology that could be extended to any GPCR. *ii)* Use of these models for their latter application in structure-based ligand design methodologies. *iii)* Combination of these structure-based methods with ligand-based techniques. *iv)* Appropriate integration with experimental studies, mainly through collaborations with medicinal chemistry and pharmacology groups, in the interdisciplinary framework typical of ligand-design projects.

- *Study of the dynamics and conformational equilibrium of GPCRs.*

With this in mind, we pursued the following objectives: *i)* The development of a specific protocol for the molecular dynamics simulations of GPCRs. *ii)* The application of MD simulations to study the conformational equilibrium of selected GPCRs, as well as to *iii)* the characterization of the receptor dimerization process.

- *Integration of the homology modelling and MD simulation protocols, developed in the previous milestones, into an automatic pipeline that is offered as a web service open to the scientific community.*

This milestone will provide important benefits: *i)* to our research group, increasing its competitiveness in the field of GPCR structural biology, *ii)* and to the scientific community, which will have a free and easy access methodology without needing prior computational modelling expertise, enabling an exhaustive structural and dynamic characterization of this pharmaceutically relevant superfamily of receptors.

4. RESULTS AND DISCUSSION

Provided a general introduction of the biology and pharmacology of the considered GPCRs, and once detailed the computational techniques employed to achieve the main objectives of this thesis, the most remarkable results are presented below. These results are mainly reported as a series of scientific publications, reproduced in *Section 7*, and are here discussed and put in context of the objectives depicted in the previous section.

- *Development and application of a robust homology modelling methodology in GPCR computer-aided ligand discovery and design.*

As described in detail in *Section 2.1.1*, we have employed a homology modelling protocol in order to obtain high-quality structural models of GPCR structures. This general protocol has been used in several computational drug discovery and design efforts, including the combination of SB and LB approaches, as summarized below.

The first computer-aided ligand discovery study is described in Paper **I**, and corresponds to an interdisciplinary project carried out in collaboration with the group of Pharmacology of Prof. Loza (University of Santiago de Compostela, USC) and the Chemistry Department of University of Minho (Prof. Proença). Here we assisted in the discovery of novel potential antipsychotic scaffolds from the original chemical library of U. do Minho, previously uncharacterized in any GPCR. We computationally screened this library of 1622 compounds with the shape-based VS methodology from Openeye suite of programs. This approach was selected due to its ability for retrieving hits with novel chemistries (181), and demonstrated good performance in retrospective VS on aminergic receptors (183). In fact, seven pharmacological hits were discovered, using the reference APC clozapine as query for the screening. The *in vitro* affinity of the identified hits identified was evaluated in the human 5-HT_{2A} receptor and showed a K_i in the micromolar range. Interestingly, these hits present a novel 9-methyl-purine scaffold with either piperazine or piperidine rings in the exocyclic N6, and different aryl substitutions in position C2. In order to rationalize the results of this VS, a homology model of the 5-HT_{2A} receptor was built using the structure of β 2ADR as template (50), employing for the first time the protocol for GPCR modelling described in *Section 2.1.1*. Interestingly, both the shape-based VS and the docking results agreed in their proposed biological superposition of the most active hit and the reference molecule, clozapine. The binding modes of the two compounds share the conserved salt bridge between Asp^{3.32} of the receptor and the positively-charged nitrogen of the ligands, in agreement with previous computational studies (215), and provided hints on the possible 10-fold lower affinity of the hit compounds, as compared to clozapine. Specifically, the purine-based ligand lacks the interaction with Thr^{3.37} predicted for clozapine, and instead establishes a hydrogen bond with Ser^{5.46} that leads to a steric clash of a methyl group of the 9-methyl-purine scaffold with TM5.

In order to further assess the homology modelling and ligand docking pipelines proposed in the previous work, our laboratory took part on the GPCR Dock 2008 assessment (see *Section 2.1* and Paper **II**). Here, our group performed blind predictions of the structure of the complex of A_{2A}AR with the potent antagonist ZM241385 (group code 5800 in the competition) right before the release of the crystallographic structure. After a template selection process, we generated and reported several homology models of the target receptor with MODELLER, using β 2ADR as template, followed by a model assessment on the basis of their stereochemical quality with Procheck, and refined with the aid

of the Molprobiy server. We submitted 5 different complexes, selected and internally ranked on the basis of the mutagenesis data available for antagonist binding on A_{2A}AR (26). The results of the contest, including all the structural models, are open to the scientific community and available at http://jicimpt.scripps.edu/gpcr_dock.html. The results of our research group are summarized in Fig. 14. Our complexes provide over-the-average predictions within the total number of submissions coming from 29 different research groups. In terms of predicting the ligand binding mode, our best complex reproduced 11 ligand-receptor native contacts and obtained a combined Z-score of 0.31,² being ranked in position 43th within the total of 208 computationally-predicted complexes. In terms of protein homology modelling, our A_{2A}AR models showed low root-mean square deviation (RMSD) for the trace of C α atoms compared to the experimental structure (ranging from 3.3 to 3.8 Å depending on the model), being ranked as the top-15 homology model. From this analysis, it becomes clear that although having presented a high quality homology model, we did not perform as well in the prediction of the binding mode of ZM241385. Indeed, in our model of the complex, the ligand presented a flipped orientation compared to the crystallographic pose, undoubtedly influenced by the relative symmetry of the molecule around the exocyclic nitrogen. In fact, a latter review of the pool of docking solutions provided by GOLD revealed low-ranked binding modes we discarded for further consideration, which were much closer to the experimental one (see Fig. 14C). The lesson learned was that, on top of the ranking provided by automated scoring functions, the researcher must properly incorporate all the available experimental data into the molecular modelling process. This is also a general outcome of the two editions of the GPCR Dock competitions (in 2008 and 2010), where the most accurate models were produced by researchers with proven knowledge of the biochemistry and pharmacology of the considered receptors, as are the cases of Costanzi in ARs (137), Selent *et al.* in aminergic and de Graaf and colleagues in peptide-binding receptors (138).

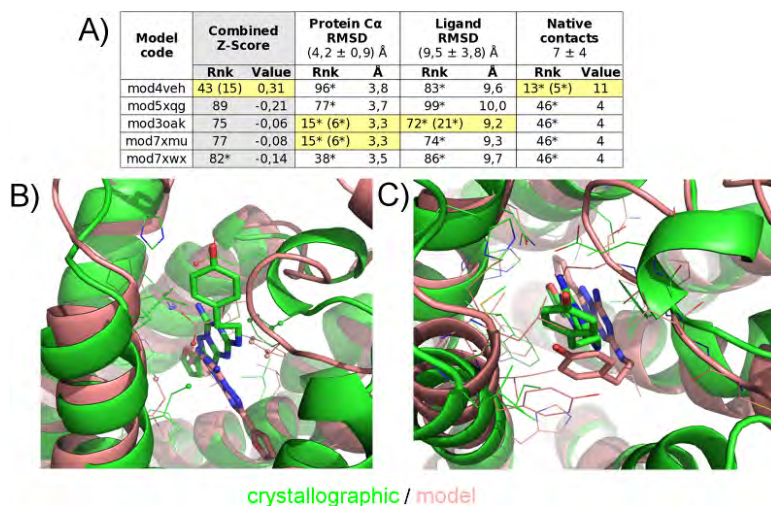


Figure 14. Summary of the participation of our research group in the GPCR Dock competition (code 5800). **(A)** Statistics of the submitted models, following their ranking on several metrics respect the crystallographic structure. Average values of all submitted models are indicated in brackets in the first row. The rankings are referred to the individual models, numbers in brackets indicate the rank only taking into account the best solution of each research group, and asterisks denote degenerated ranking. **(B)** Best submitted complex solution (model code mod4veh) **(C)** Discarded solution from the final selection stage, significantly resembling the crystallographic pose.

² The Z-score is a dimensionless measure that normalizes a value respect to the mean and standard deviation of the sample, in this case considering the RMSD and native contacts of the ligand.

In a follow-up study, we generated homology models of the remaining human ARs subtypes: A_1 , A_{2B} and A_3 ARs, using the recent structure of A_{2A} AR in inactive-like conformation (56). Despite the high sequence identity within the family — A_{2A} AR presents a TM sequence identity between 52% and 72% with the rest of subtypes—, there are remarkable differences in ligand binding affinity among ARs. In order to investigate these selectivity issues, we generated a pseudo-sequence alignment of the binding site residues of ARs subtypes, in the line of previous chemogenomics efforts (9, 10), and as also attempted by other authors (109, 216). Such analysis, illustrated in Fig. 15A, allowed us to detect the most promising residue substitutions to explain ligand selectivities. We selected three antagonists with different chemical scaffolds and selectivity profiles in the ARs family to conduct a docking exploration on the generated homology models: the triazolo-quinazoline CGS15943, the xantine derivative DPCPX and the triazolo-triazine ZM241385 which had been co-crystallized with A_{2A} AR (see Fig. 15B). The biological superimposition proposed for CGS15943 and DPCPX in the A_{2A} AR structure, together with the co-crystallographic ligand (successfully re-docked as a test case), is in agreement with previous ligand-based studies (217). This was further corroborated for the case of DPCPX, where our predicted binding mode is in agreement with the recently solved structure of A_{2A} AR with the xantine derivate XAC (17), shown in Fig. 15C. The same superimposition is observed for A_{2B} AR, except subtle differences that might be probably attributed to the presence of Val^{6.51} in that receptor instead of the conserved Leu in the rest of the ARs. Regarding other selectivity issues that might be explained by this docking study, the ligand ZM241385 frequently presents a flipped pose in the A_1 and A_3 AR as compared to the binding mode in the crystal structure with the A_{2A} AR: these observations are in agreement with a 10- to 70-fold lower affinity of this ligand in A_1 AR and A_3 AR compared to both A_2 subtypes. Also, the conserved solution of DPCPX in A_3 AR can be explained by the less voluminous Leu in position 7.35 in this receptor, as compared to the otherwise conserved Met^{7.35} in the family. Accordingly, this compound presents 50- to 1000-fold less affinity in the A_3 AR as compared to other subtypes.

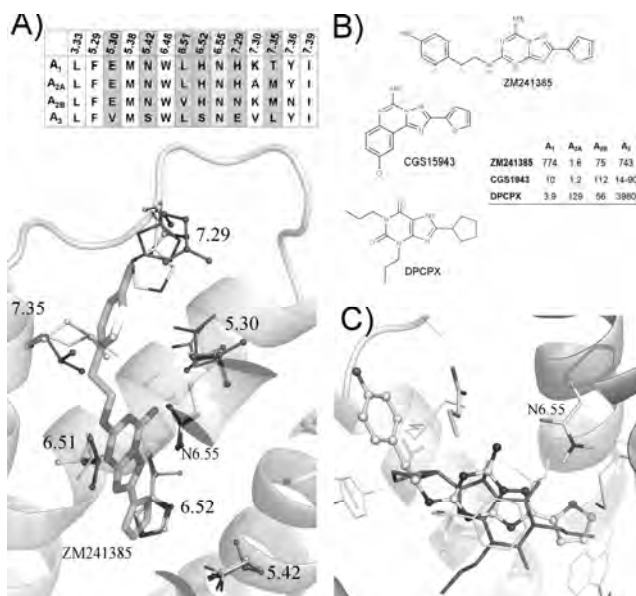


Figure 15. Systematic homology modelling and ligand docking for studying selectivity issues on ARs. **(A)** The pseudo-sequence alignment of ARs residues at 5A of the crystallographic ligand as observed in PDB structure 3EML, together with the representation of key positions with different mutations in the rest of ARs subtypes. **(B)** Structures and K_i affinities (nM) of the selected antagonists for the docking study on the aforementioned structures. **(C)** Biological superimposition of DPCPX (dark sticks) after docking on 3EML structure (ZM241385 in white ball & sticks), in agreement with A_{2A} in complex with the analogous XAC. See Annex I for details.

On the basis of this modelling study of the members of ARs family, and in collaboration with the medicinal chemistry group of Prof. Sotelo (USC), our laboratory contributed to the computational design and the rationalization of the pharmacological profile of a novel library of diaryl-pyrimidine derivatives with potent and selective profiles on A₃AR. This study is presented in Paper **III** of this thesis. Here, a combination of SB and LB computational approaches was considered, using the homology model of the human A₃AR developed in our laboratory (described above) as starting point. A systematic docking of the compounds of the chemical library allowed the determination of the bioactive conformations, which were used as input for a 3D-QSAR modelling that was ultimately used in the assessment and the growth of the chemical library. In the first stage, a consensus binding mode was determined for the compounds of the chemical library. Accordingly, a double hydrogen bond between the key residue Asn^{6.55} with both the exocyclic amido group and the closest nitrogen in the pyrimidine scaffold in all ligands was observed, together with a π -stacking interaction with Phe^{5.29} at EL2 (both residues totally conserved in ARs), and van der Waals interactions with Leu^{6.51}. Notably, the recent crystal structure of A_{2A}AR in complex with a 1,2,4-triazine derivative (55), a scaffold structurally similar to the aforementioned pyrimidine derivatives, supports this hypothesized binding mode (see Fig. 16). The ligands exploit 3 lipophilic subsites of the receptor with chemical groups attached to their exocyclic amino/amido (L1) and aryl substituents (L2, L3). The biological molecular alignment obtained from the consensus docking pose was used as a starting point for a 3D-QSAR model with the second generation of GRid INdependent Descriptors (GRIND-2) (188) implemented in the software Pentacle (218). The resultant 3D-QSAR model presented a satisfactory statistical quality, with a fit of $r^2=0.86$, and a predictive ability of $q^2=0.67$ obtained by LOO-CV. The main pharmacophoric properties of the series are represented by corresponding correlograms, i.e., alignment-independent vectors of relevant MIF node pairs extracted from the model (see Section 2.2.2.2 for details). Most remarkably, these correlograms suggests a relationship between the size of substituents at L1 and L2/L3 subsites (the bulkier in L1, the smaller in L2/L3, and *vice-versa*) in order to achieve increased activity on A₃AR. An advantage of this combination of SB and LB methodologies is that the 3D-QSAR descriptions are back-projectable to the binding site. Thus, hydrophobic interactions at the L1 site (extracellular tip of the binding site) are detected by O-TIP and DRY-TIP cross-correlograms, reflecting interactions with Ile^{6.58} and Leu^{7.35}, and explaining the reduced affinity of aminopyrimidine derivatives due to the lack of substitutions at that subsite. Also, the reduced potency of 2-amido-pyrimidines respect to the 4-amido congeners is represented by O-N1 and N1-N1 correlograms, also suggesting the presence of a structural water molecule mediating ligand-receptor interactions for the latter derivatives. Additionally, selectivity issues can be easily detected by the sequence analyses of binding sites of ARs described above. In this regard, positions 5.30, 5.42 and 6.52, together with the already mentioned 7.35, have been pointed out as the mostly relevant for selectivity issues of these ligands in the context of ARs. Specifically, His^{6.52} and Asn^{5.42} in A₁, A_{2A} and A_{2B} ARs are substituted by a serine residue in both positions of A₃AR, allowing bulkier substitutions of the ligands in this subsite (L3). Additionally, hydrophobic and more voluminous substitutions at the exocyclic amido group are favoured in A₃AR due to the presence of Val^{5.30} at EL2 (where a Glu^{5.30} appears for the rest of subtypes) and a smaller Leu^{7.35} (compared to Met in A₂ subtypes). The described computational pipeline was employed in the design of a set of 6 novel ligands, with bulky substituents in the amido group of the 2,6-diaryl-4-amidopyrimidine scaffold with excellent results (see Paper **III**), and is being currently employed in the design on novel ligands with enhanced pharmacological profiles towards A₃AR.

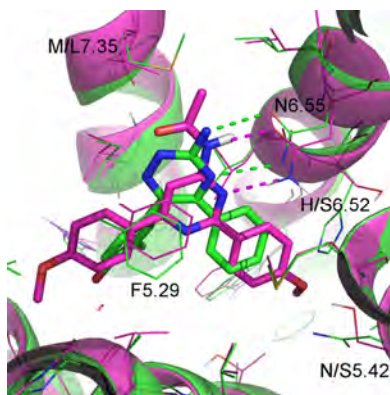


Figure 16. Comparison of the computationally-predicted binding mode of ISVY130 in A_3AR (blue), and the crystallographic structure of a 1,2,4-triazine derivative in complex with $A_{2A}AR$ (PDB code 3UZC, in magenta). The double hydrogen bond with Asn^{6.55}, together with key binding site residue residues, are indicated. Key residue substitutions in A_3AR provide a broader sub-pocket between TM5 and TM6.

Following the structural characterization of ARs discussed above, we aimed to combine these studies with ligand-based approaches for the discovery of novel potent and selective ligands towards the different ARs. This was the subject of a research stay in the laboratory of Dr. Andreas Bender, at the Unilever Centre (University of Cambridge, UK) which main results are outlined in Annex II. In this work, we screened the chemical library of the Galician Network for Drug Discover, which counts with the participation of our group and is managed by the laboratory of Prof. Loza (USC). A VS protocol was designed and executed as follows:

i) Training bioactivity models for each ARs subtype (A_1 , A_{2A} , A_{2B} , A_3) with the tool MOLPRINT2D (see Section 2.2.2.3), with data extracted and curated from the ChEMBL database (<https://www.ebi.ac.uk/chembl/>). Molecules with a pK_i above 7 were considered as active, while those with pK_i below 6 were categorized as inactive. Since most of the molecules deposited in this dataset come from successive medicinal chemistry campaigns that explore series of analogues, we removed this redundancy by means of clustering with JKlustor (ChemAxon, using ECFP4 fingerprints) in order to maximize the molecular diversity of the training sets. The datasets of diverse active and inactive ligands for each subtype, with an approximate population of 200 compounds each, were complemented with 800 random compounds from ChEMBL added to the inactive subset in order increase the resolution of the scoring. After obtaining the ligand score for each bioactivity model, we set a multi-objective selection criteria: a compound would have to be ranked in the top 10% of one subtype and below that threshold in the rest of subtypes for being a promising selective compound. An additional less restrictive criteria, termed as “promiscuous”, was selected for fewer ligands within the top10% in at least two out of the four possible AR models.

ii) A post-filtering stage on the basis of physicochemical properties and (dis)similarity with known actives was performed in order to retain compounds with drug-like properties and novel scaffolds.

iii) The final selection of compounds to be experimentally tested on the four ARs was assessed by docking simulations on our 3D models of ARs, paying special attention to interactions with the key residue Asn^{6.55}. Finally, 40 compounds were assayed on all subtypes, which generally present novel scaffolds in the medicinal chemistry of ARs (i.e., single-ringed cores and extended side chains). From these ligands, we found 11 compounds over the minimum threshold of 30% inhibition (at 10^{-5} M concentration) on at least one receptor subtype, with 2 of them above 50% in A_1AR and $A_{2B}AR$ (see Fig. 17). The performance of the A_3AR and promiscuous models are especially remarkable. The present results show that the LBVS methods are able to obtain high hit rates —

RESULTS AND DISCUSSION

especially considering the nature of the screened chemical library—, in agreement with previous published screenings on the A_{2A} AR (219). This has also been one of the conclusions of a recent cross-screen on three GPCRs (220), where most importantly the prediction of selectivity appears to be still a difficult issue for the state-of-the-art techniques. This study is on a second phase at our research group at the time of the writing of this thesis, consisting on a search of analogues from vendors with prospectively enhanced properties in order to provide hits with higher affinities.

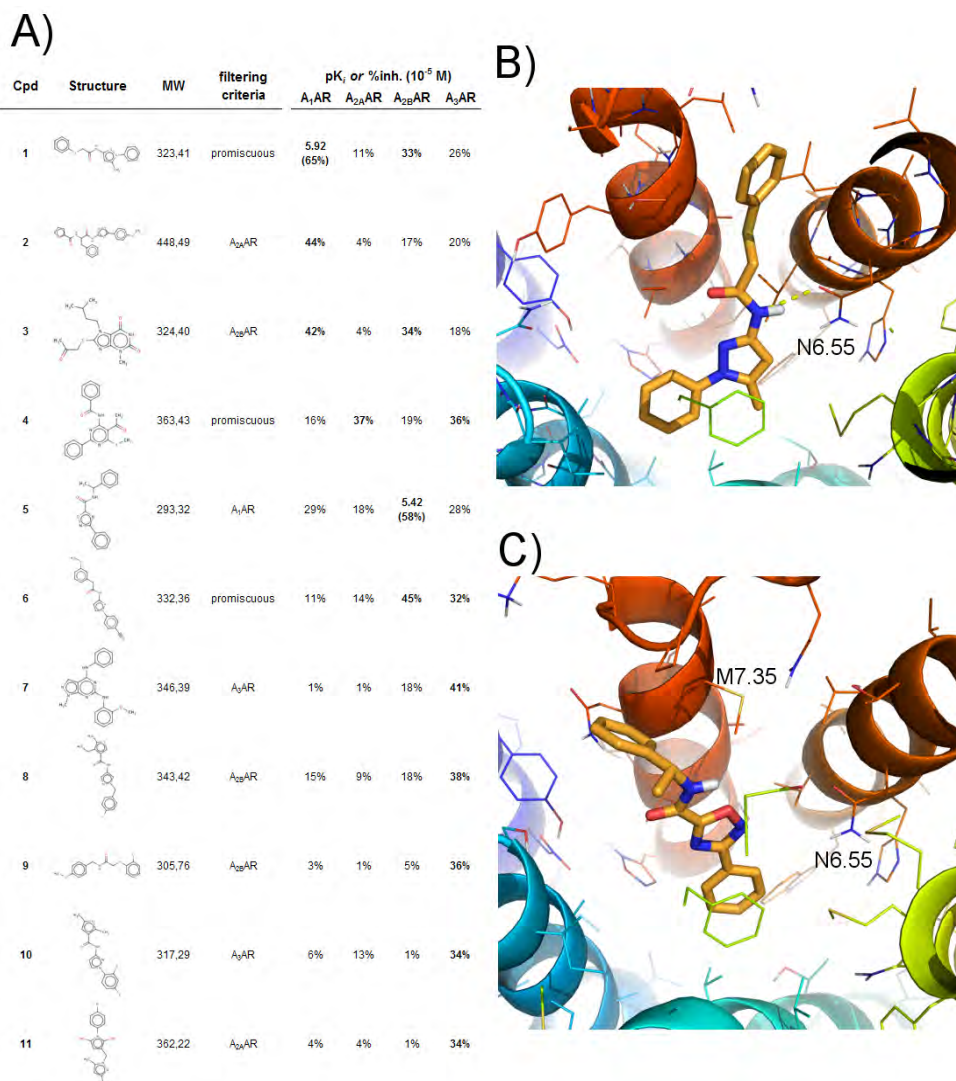


Figure 17. Pharmacological hits obtained in an ongoing campaign for discovering novel and selective ligands on ARs. **(A)** Discovered pharmacological hits, together with their molecular weight, the selection criteria met in the VS, and their pharmacological data on ARs. On the right hand side, binding modes of compound **1** in A₁AR **(B)** and compound **5** on A_{2B}AR **(C)** In the latter binding mode, no significant interactions were predicted for compound **5** in A_{2B}AR (in fact, was selected for A₁AR), precluded by the rotamer of Met^{7.35}.

- *Study of the dynamics and conformational equilibrium of GPCRs.*

In the recent years, several studies of the dynamics of membrane receptors have been performed through all-atom MD simulations. This has been possible due to the parallel advances in structural biology of GPCRs as well as important developments in computational software and hardware. In this scenario, our research group, in collaboration with Dr. Piñeiro and Prof. Brocos (Department of Applied Physics, USC), has developed of a computational protocol for the automated setup and performance of MD simulations for the particular case of GPCRs. The protocol, which is deeply described in *Section 2.3.3* and originally published in Paper **IV**, considers an atomistic model of the biological membrane, consisting of a POPC bilayer, solvated with explicit water molecules and neutralized with counter-ions, under the PBC scheme implemented in the software package GROMACS (221). Using this pipeline we performed and analysed MD simulations of the X-Ray structure of A_{2A} AR in its inactive-like conformation (PDB code 3EML) (56), as well as of a homology-derived model of the related A_{2B} AR subtype, as reported in Paper **IV**. The use of simulation replicas (identical starting points with different random seeds for the assignment of initial velocities) and several setups (absence or presence of crystallographic ligands, consideration of crystallographic water molecules, and residues with different protonation states) were considered. After exhaustive equilibrations, the production phases of the main simulations spanned for 100 ns, an appropriate timescale for characterizing initial conformational rearrangements from the inactive-like state towards more active-like forms of the receptors. Additionally, the use of simulation replicas provides a better thermodynamic sampling, increasing the statistical significance of the observed events.

One of the microswitches studied in these simulations was the *ionic lock*, a salt bridge established in inactive receptor conformations between residues Arg^{3.50} and Glu^{6.30} (in the intracellular tips of TM3 and TM6, see *Section 1.2.1*). This interaction, however, was absent in the experimental inactive structure of the A_{2A} AR (56), and consequently in the homology-derived model of the A_{2B} AR. However, we observed a quick and stable formation of this lock in the different MD simulations replicas on both considered receptors. Moreover, we could characterize the tight hydrogen bond network that supports this “lock”, which includes Asp^{3.49} and Tyr^{3.60(IL2)}. This is in agreement with previous simulations of A_{2A} AR (222) and with computational studies of inactive-like crystal structures of β -adrenergic receptors with initially broken ionic locks (50, 52), where simulations around the microsecond scale showed a spontaneous and generally stable formation of this interaction (223-227). A common feature of the crystallization process of both A_{2A} AR and β -adrenergic receptors is the use of the T4L protein fusion strategy, which was lately discussed as the source of distortions in the intracellular region and probably responsible of the absence of this salt bridge (83). In fact, in the X-Ray structures of A_{2A} AR in complex with inverse agonists, which were obtained using mutation-stabilized receptors, the ionic lock is formed (57). This result agrees with our simulations and exemplifies the predictive ability of MD in the refinement of crystal structures in a model resembling the physiological environment of the receptor. As a final note, the reformation of the ionic lock has been recently observed in MD simulations of active-like conformations of β 2ADR in the microsecond scale (228).

The second examined microswitch was the rotameric transition of the side chain of Trp6.48 from *gauche+* to *trans*, known as the *toggle switch*, and largely associated to the activation mechanism of GPCRs (93-95). This event was observed on both A_{2A} AR and A_{2B} AR, where a cluster of conserved residues in ARs —namely Asn^{5.42}, Gln^{3.37}, Cys^{5.46} and His^{6.52}— was found to co-operate with the formation of this switch through concerted conformational changes in order to form a novel hydrogen bond network. While this rotameric transition was not observed in the only simulation of β 2ADR crystal structure that explicitly analyzed this switch (227), it was indeed reported in large-scale MD simulations on a homology model of the dopamine D_2 receptor, in this case considering the

allosteric modulation of sodium ions located in the vicinity of this region (229). Very recently, this sodium allosteric binding site, largely proposed for many GPCRs, was experimentally observed in a crystal structure of the A_{2A}AR and is located in the tight hydrogen bond network of GPCR conserved residues at TM1-TM2-TM6-TM7. Interestingly, in our simulations the solvent occupancy of this region in A_{2A}AR is well reproduced, regardless of the consideration of crystallographic waters in the MD setup. Currently, we are studying the sensitivity of this receptor to the presence of sodium ions and other allosteric modulators, using a similar computational methodology, in collaboration with the groups of Prof. Stevens (Scripps Research Institute, California) and Prof. IJzerman (Leiden University).

Going back to Paper **IV**, we were also interested in the structural role of two other particular motifs in ARs that involve key histidine residues. The first one refers to the interaction between residues of EL2 (Glu^{5.30}) and EL3 (position 7.29), closing the top of the binding site. A tighter interaction in A_{2A}AR, especially considering the more plausible charged state of His^{7.29}, is observed as compared to the A_{2B}AR where this histidine is replaced by the neutral Asn^{7.29}. This aminoacidic substitution is responsible for a higher mobility of the extracellular loops of the latter receptor according to our simulations, suggesting a higher diffusion rate of compounds in this receptor that might explain the lower affinity for most ARs ligands on A_{2B}AR. The second motif involves the totally AR conserved residues Glu^{1.39} and His^{7.43}. Here, the initial hydrogen bond suggested by the crystal structure is broken along the simulations, significantly regulating the dynamics of TM1 and TM7. The tautomeric state of His^{7.43} protonated in delta, as modelled by us, is in agreement with latter crystallographic structures of A_{2A}AR in complex with agonists (53, 54). The exhaustive pK_a calculations performed suggested a charged state for this histidine in the A_{2A}AR, a possibility that was computationally assessed together with the *in silico* proposal of a His^{7.43}Lys mutant. Our results propose the existence of a salt bridge between residues Glu^{1.39} and His^{7.43}, and issue that remains to be confirmed by molecular biology and pharmacology experiments. Finally, a comparison between the lately crystallized structure of A_{2A}AR in an active-like form (54) and the endpoint of one of A_{2A}AR simulation replicas reveals a partial agreement of the conformations of the extracellular region of the receptor (see Supporting Information of Paper **IV**). In fact, the induced-fit promoted by the co-crystallized agonist UK-432097 results in a disruption of the interaction between EL2 and EL3, as well as a conformational change of the extracellular half of TM7 outwards the binding crevice, both events being well reproduced in our simulations. The experimental confirmation of our results is encouraging, although we acknowledge that the more drastic conformational changes that drive the receptor into its active state could not even be achieved by our computational experiments, due to the timescale affordable by our simulation approach (equilibrium-MD).

The computational protocol presented in Paper **IV** was adapted in order to analyse the phenomenon of receptor dimerization in Paper **VI**. Provided the novel experimental evidences of GPCR dimerization as revealed by the crystallographic structures of CXCR4 (65), released in October 2010, we decided to further investigate the structural determinants of these protein-protein interactions. The CXCR4 homodimers show a previously unexpected dimerization mode involving TM5 and TM6 (see Section 1.2.3 and Paper **IV** for details), observed in two slightly different crystallographic dimers: one co-crystallized with the organic ligand 1T1t (PDB code 3ODU); and the other —assigned by the authors— co-crystallized with the cyclic peptide CVX15 (PDB code 3OE0). We performed extensive MD simulations of the apo forms of these two molecular models, considering both the stabilized mutant crystallographic form, as well as the wild-type forms of the receptors. In this respect, our MD simulations showed punctual distortions on the wild-type protomers, thus justifying the observed increased thermal stability of the mutant forms. Generally speaking, the simulations support the dimerization mode observed in both structures, since the position of the protomers remained stable on the initial pose, although certain rearrangements in the relative orientation of protomers were observed: the buried surface area of the dimerization interface growth in all cases, mainly driven by

hydrophobic interactions between the receptors (keeping those observed in the crystallographic snapshots). The conformational changes were more significant for the case of 3ODU dimer, due to the lower initial contacts in the intracellular half of the dimerization interface compared to 3OE0 structure (65). This behaviour is also line with the observations reported in the recent structure of μ -OR (71), which adopts a higher-order oligomeric arrangement with commonalities with the dimerization mode observed for CXCR4 structures. The μ -OR structure further supports the relevance of the TM5-TM6 interface (see Fig. 11). However, the relatively high buried surface area between the intracellular T4L fusion proteins in 3ODU protomers —unlike 3OE0 or μ -OR structures— might induce a bias in its initial conformation justifying the observed conformational changes. Given the transient nature of GPCR dimers (230), this computational evaluation significantly explores expectable structural rearrangements from the crystallographic structures of CXCR4, supporting the ligand-independent dimerization of this receptor, and concluding that the interface of 3OE0 structure presents a reliable high-resolution depiction of GPCR dimerization.

Additionally, we also studied the dynamics of key residues of CXCR4 involved in ligand binding and in the process of HIV infection. We focused on differences of the tautomeric state of His^{3,29} according to the conformations of surrounding residues, characterizing the influenced of the co-crystallographic ligands in that region. Thus, while the initial salt bridge between Asp^{4,60} and Arg^{5,27(EL2)} remained stable for the apo form of the IT1t-bound structure (3ODU), a strong hydrogen bond between His^{3,29} and Asp^{4,60} occurs instead in the apo version of the CVX15-bound structure (3OE0). We suggest that specific induce-fit effects occur upon the binding of the cyclic peptide, a conclusion which is further supported by the available mutagenesis data. Our work concludes that the conformation of 3ODU structure is more appropriate for the structure-based drug design of small ligands.

As a final note, despite that the simulations performed in Papers **IV**, **VI** do not reach the microsecond timescale —as opposed to other contemporary simulations of GPCRs— the range of hundreds of nanoseconds is considered as appropriate taking into account: *i*) the short timescale of the dynamic and structural events that were aimed (initial conformational rearrangements from an inactive conformation, or the stability of GPCR crystallographic dimers), *ii*) the use of simulation replicas to increase statistical sampling and *iii*) the available computational resources. We can conclude that our methodology is in the state-of-the-art of statistical dynamic sampling techniques, and at the same time provide several methodological advances for the MD simulation of GPCRs.

- *Integration of the homology modelling and MD simulation protocols, developed in the previous milestones, into an automatic pipeline that is offered as a web service open to the scientific community.*

As discussed before, the structures of many GPCRs remain experimentally unsolved despite the recent advances in their crystallization. Following the previously described methodologies on GPCR homology modelling and MD simulations, developed through this thesis with successful results, we considered the possibility of automating the process in order to access to a higher coverage of the structural knowledge of GPCRs. Several services for this purpose are available on the web, which we introduced in *Section 2.1.2* and reviewed in Paper V, at the same time that we presented our GPCR-ModSim webserver (<http://gpcr.usc.es>). We here discuss the main advances of this pipeline:

i) The first critical stage of homology modelling is the selection of the most appropriate template for the target receptor. For this purpose we have elaborated a structural alignment of all members of GPCRs with solved structures, and divided them according to their functional conformation (either using the 14 inactive- or the 5 active-like receptors available to date), see Fig. 18. The last update of our structural alignment was performed with the aid of the software Strap (<http://3d-alignment.eu>), which produces a multiple-sequence alignment (MSeqA) derived from a multiple-structure alignment (MStA) with the TM-align algorithm (231). The resulting MSeqA was manually reviewed, paying special attention to gaps and indels within the TM region due to local distortions, and the correct alignment of the two cysteines involved in the conserved disulfide bridge (TM3-EL2), among other aspects. This refined MSeqA of the templates is stored in the server as the profile to which the sequence of the query receptor is aligned to. In this sense, it is worth noting that the query sequence can be retrieved directly from its UniprotID, or may consist of a “customized” protein mutant. The initial query-templates alignment is automatically made with Clustal0 (232), but can be further refined by the user with a JalView applet. The server automatically provides statistics of sequence identity for all the different regions of the receptor (individual loops and TMs), and by default suggests the template with the highest sequence identity in the TM region with the target receptor. A practical application of the template selection method offered by the server was performed in the homology modelling of neuropeptide Y2 receptor (233).

ii) Once the query-template sequence alignment is configured, the user is allowed to perform up to 10 models with MODELLER 9v8 (155), which are ranked according to the statistical potential DOPEHR (234). In order to further select the models, stereochemical quality reports generated by Molprobit (158) are provided, together with a JalView applet that allows a quick visual inspection. A Python script for a local visualization of all the models in PyMOL can be downloaded as well. On top of this initial modelling, the user is allowed to run up to 5 loop optimizations in selected model, as implemented in the loopmodel routine (156) with the same tools for selection and visualization as described above. The introduction of custom disulfide bridges can be performed in either initial and/or loop modelling stages.

iii) Finally, one of the main novelties of our web service is the setup and MD equilibration of the receptor in an atomistic model of the biological membrane, following the protocol developed in Paper IV. The user can perform this equilibration in one of the generated models from the server, or even upload an external PDB structure of a GPCR, which will be conveniently inserted in the simulation box and equilibrated automatically by the server. After 2.5 ns equilibration, the user will be provided the refined system (with the corresponding MD trajectory), a brief report and all necessary files for a local extension of the simulation with the GNU-licensed software GROMACS v4.0.5 (221).

Right after the crystallization of human histamine H₁ receptor in complex with doxepin (67),

we used this novel structure to carry out a comparative analysis of the performance of three web-based servers for the modelling of GPCRs, namely GPCR-ITASSER (164), GPCR-SSFE (148) and our GPCR-ModSim service, using default parameters and minimum user intervention. All the three tools were able to obtain good quality models, where the RMSD of C α atoms of the TM bundle was below 2.0 Å. The accuracy achieved by GPCR-ITASSER and GPCR-ModSim was especially remarkable, including the loop regions (C α RMSD for the whole receptor of 2.09 and 2.31 Å respectively, see Paper VI). It is important to note that the protocol implemented by GPCR-ModSim allows the generation of such models in minutes, while the exhaustive threading and loop refinement method of GPCR-ITASSER typically takes several days. A second advantage of our method, outlined in Fig. 18, is that it is the only service that allows an automated MD exploration of the model, which is a recognized step forward in the refinement and sampling of the receptor structure(s) for either drug design (235) or structural biology (228) applications.

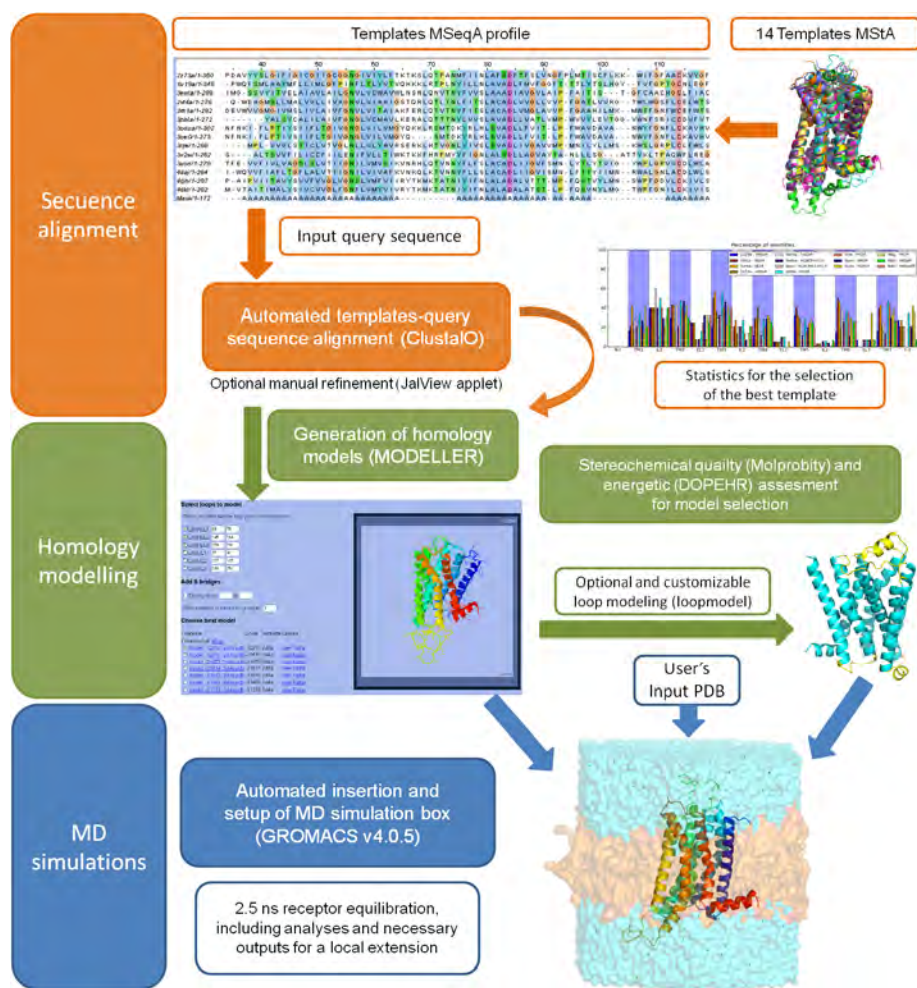


Figure 18. Outline of the computational protocol implemented in GPCR-ModSim server. The main interactions (input/output) with the users in the three stages of the pipeline (sequence alignment of the query receptor, homology modelling and MD simulations) are indicated.

As a conclusion, the GPCR-ModSim web server provides the most complete pipeline of GPCR homology modelling to our knowledge, including the performance of MD simulations for the first time. A minimum user intervention is necessary to obtain reliable homology models with state-of-the-art techniques, although extensive flexibility is considered in order to allow proper decision-making by the user in the key steps, such as the refinement of the target-template sequence alignment or the model selection. Users without a strong computational background can find this service useful for tasks such as the spatial location and surroundings of key residues (including protein mutants), the assessment of the stability of protein mutants, or the study of selectivity issues within a family of receptors (as we exemplified for ARs). Continuous updates are performed in the server, the most notable being: *i*) the incorporation of any novel GPCR-crystal structure, *ii*) the consideration of receptor dimers, *iii*) the inclusion of ligands, either allosteric and/or orthosteric, or cofactors (i.e. specific cholesterol or lipid molecules) in the MD protocol, to account for the complexity of the receptor system suggested by recent evidences (64, 101).

5. CONCLUSIONS

State-of-the-art techniques in the area of molecular modelling and computer-aided drug discovery and design have been used as a basis for further methodological developments, as well as in applied studies in the biochemistry, medicinal chemistry and pharmacology of selected members of GPCRs. The main achievements of this thesis include:

- The development of a complete and automated computational pipeline for the homology modelling, structural refinement and dynamic analysis of GPCRs, which has been tested in different systems such as a comparative analysis of adenosine receptors or the homodimerization of CXCR4. We have fully automated this pipeline in the GPCR-ModSim web server (<http://gpcr.usc.es>), which can be used by researchers worldwide without prior modelling knowledge.
- A structural characterization of the conformational equilibrium and ligand selectivity on the Adenosine Receptors family, by a proper combination of the first experimental structure of the A_{2A}AR in complex with a potent antagonist (ZM241385), the available mutagenesis data and our computational methodology. This provides a map of the functional and binding selectivity issues in this family of receptors, which is here used in drug design applications.
- The successful exploration of combinations of Structure-Based and Ligand-Based approaches in GPCR drug discovery and design. The use of homology models together with techniques such as Shape-Based Virtual Screening and 3D-QSAR allowed the discovery and rationale of the binding of novel scaffolds for future antipsychotic compounds from an original chemical library, and the design of a novel series of diarylpyrimidines with a potent and selective pharmacological profile on the A₃AR.

6. REFERENCES

- (1) Paul, S. M., Mytelka, D. S., Dunwiddie, C. T., Persinger, C. C., Munos, B. H., Lindborg, S. R., and Schacht, A. L. (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9, 203-14.
- (2) Jorgensen, W. L. (2004) The many roles of computation in drug discovery. *Science* 303, 1813-1818.
- (3) Strachan, R. T., Ferrara, G., and Roth, B. L. (2006) Screening the receptorome: an efficient approach for drug discovery and target validation. *Drug Discov. Today* 11, 708-716.
- (4) Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009) The structure and function of G-protein-coupled receptors. *Nature* 459, 356-363.
- (5) Overington, J. P., Al-Lazikani, B., and Hopkins, A. L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993-6.
- (6) Gutkind, J. S. (1998) The pathways connecting G protein-coupled receptors to the nucleus through divergent mitogen-activated protein kinase cascades. *J. Biol. Chem.* 273, 1839-1842.
- (7) Kenakin, T. (2004) Principles: receptor theory in pharmacology. *Trends Pharmacol. Sci.* 25, 186-192.
- (8) Mirzadegan, T., Benko, G., Filipek, S., and Palczewski, K. (2003) Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin. *Biochemistry* 42, 2759-2767.
- (9) Surgand, J. S., Rodrigo, J., Kellenberger, E., and Rognan, D. (2006) A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* 62, 509-38.
- (10) Gloriam, D. E., Foord, S. M., Blaney, F. E., and Garland, S. L. (2009) Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. *J. Med. Chem.* 52, 4429-42.
- (11) Fredriksson, R., Lagerström, M. C., Lundin, L.-G., and Schiöth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256-1272.
- (12) Mueser, K. T., and McGurk, S. R. (2004) Schizophrenia. *Lancet* 363, 2063-72.
- (13) Schultz, S. H., North, S. W., and Shields, C. G. (2007) Schizophrenia: a review. *Am Fam Physician* 75, 1821-9.
- (14) Creese, I., Burt, D. R., and Snyder, S. H. (1976) Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. *Science* 192, 481-3.
- (15) Meltzer, H. Y., Matsubara, S., and Lee, J. C. (1989) Classification of typical and atypical antipsychotic drugs on the basis of dopamine D-1, D-2 and serotonin₂ pKi values. *The Journal of Pharmacology and Experimental Therapeutics* 251, 238-246.
- (16) Hippus, H. (1999) A historical perspective of clozapine. *J. Clin. Psychiatry* 60 Suppl 12, 22-3.
- (17) Meltzer, H. Y., Alphs, L., Green, A. I., Altamura, A. C., Anand, R., Bertoldi, A., Bourgeois, M., Chouinard, G., Islam, M. Z., Kane, J., Krishnan, R., Lindenmayer, J. P., and Potkin, S. (2003) Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). *Arch. Gen. Psychiatry* 60, 82-91.
- (18) Arranz, M. J., and Kapur, S. (2008) Pharmacogenetics in psychiatry: are we ready for widespread clinical use? *Schizophr. Bull.* 34, 1130-44.

- (19) Gray, J. A., and Roth, B. L. (2007) The pipeline and future of drug development in schizophrenia. *Mol. Psychiatry* 12, 904-22.
- (20) Roth, B. L., Sheffler, D. J., and Kroeze, W. K. (2004) Magic shotguns versus magic bullets: selectively non-selective drugs for mood disorders and schizophrenia. *Nat. Rev. Drug Discov.* 3, 353-359.
- (21) Fredholm, B. B., AP, I. J., Jacobson, K. A., Linden, J., and Muller, C. E. (2011) International Union of Basic and Clinical Pharmacology. LXXXI. Nomenclature and classification of adenosine receptors--an update. *Pharmacol. Rev.* 63, 1-34.
- (22) Jacobson, K. A., and Gao, Z.-G. (2006) Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discov.* 5, 247-264.
- (23) Johnson, S. G., and Peters, S. (2010) Advances in pharmacologic stress agents: focus on regadenoson. *J Nucl Med Technol* 38, 163-71.
- (24) Mozzicato, S., Joshi, B. V., Jacobson, K. A., and Liang, B. T. (2004) Role of direct RhoA-phospholipase D1 interaction in mediating adenosine-induced protection from cardiac ischemia. *FASEB J.* 18, 406-8.
- (25) Yan, L., Burbiel, J. C., Maass, A., and Muller, C. E. (2003) Adenosine receptor agonists: from basic medicinal chemistry to clinical development. *Expert Opin Emerg Drugs* 8, 537-76.
- (26) Moro, S., Gao, Z.-G., Jacobson, K. A., and Spalluto, G. (2006) Progress in the pursuit of therapeutic adenosine receptor antagonists. *Med. Res. Rev.* 26, 131-159.
- (27) Hernán, M. A., Takkouche, B., Caamaño-Isorna, F., and Gestal-Otero, J. J. (2002) A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Ann. Neurol.* 52, 276-284.
- (28) Muller, C. E., and Jacobson, K. A. (2011) Recent developments in adenosine receptor ligands and their potential as novel drugs. *Biochim. Biophys. Acta* 1808, 1290-308.
- (29) Kalla, R. V., and Zablocki, J. (2009) Progress in the discovery of selective, high affinity A(2B) adenosine receptor antagonists as clinical candidates. *Purinergic Signaling* 5, 21-29.
- (30) Baraldi, P. G., Preti, D., Borea, P. A., and Varani, K. (2012) Medicinal chemistry of a(3) adenosine receptor modulators: pharmacological activities and therapeutic implications. *J. Med. Chem.* 55, 5676-703.
- (31) Viola, A., and Luster, A. D. (2008) Chemokines and their receptors: drug targets in immunity and inflammation. *Annu. Rev. Pharmacol. Toxicol.* 48, 171-97.
- (32) Feng, Y., Broder, C. C., Kennedy, P. E., and Berger, E. A. (1996) HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 272, 872-7.
- (33) Brelot, A., Heveker, N., Montes, M., and Alizon, M. (2000) Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J. Biol. Chem.* 275, 23736-44.
- (34) Berger, E. A., Murphy, P. M., and Farber, J. M. (1999) Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu. Rev. Immunol.* 17, 657-700.
- (35) Balkwill, F. (2004) The significance of cancer cell expression of the chemokine receptor CXCR4. *Semin. Cancer Biol.* 14, 171-9.
- (36) Choi, W. T., Duggineni, S., Xu, Y., Huang, Z., and An, J. (2011) Drug discovery research targeting the CXC chemokine receptor 4 (CXCR4). *J. Med. Chem.* 55, 977-94.
- (37) Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature*

- 381, 667-73.
- (38) Yost, R., Pasquale, T. R., and Sahloff, E. G. (2009) Maraviroc: a coreceptor CCR5 antagonist for management of HIV infection. *Am J Health Syst Pharm* 66, 715-26.
- (39) Schellekens, P. T., Tersmette, M., Roos, M. T., Keet, R. P., de Wolf, F., Coutinho, R. A., and Miedema, F. (1992) Biphasic rate of CD4+ cell count decline during progression to AIDS correlates with HIV-1 phenotype. *Aids* 6, 665-9.
- (40) Doranz, B. J., Grovit-Ferbas, K., Sharron, M. P., Mao, S. H., Goetz, M. B., Daar, E. S., Doms, R. W., and O'Brien, W. A. (1997) A small-molecule inhibitor directed against the chemokine receptor CXCR4 prevents its use as an HIV-1 coreceptor. *J. Exp. Med.* 186, 1395-400.
- (41) Choi, W. T., Duggineni, S., Xu, Y., Huang, Z., and An, J. (2011) Drug Discovery Research Targeting the CXC Chemokine Receptor 4 (CXCR4). *J. Med. Chem.*
- (42) Hatse, S., Princen, K., Gerlach, L. O., Bridger, G., Henson, G., De Clercq, E., Schwartz, T. W., and Schols, D. (2001) Mutation of Asp(171) and Asp(262) of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol. Pharmacol.* 60, 164-73.
- (43) Stone, N. D., Dunaway, S. B., Flexner, C., Tierney, C., Calandra, G. B., Becker, S., Cao, Y. J., Wiggins, I. P., Conley, J., MacFarland, R. T., Park, J. G., Lalama, C., Snyder, S., Kallungal, B., Klingman, K. L., and Hendrix, C. W. (2007) Multiple-dose escalation study of the safety, pharmacokinetics, and biologic activity of oral AMD070, a selective CXCR4 receptor inhibitor, in human subjects. *Antimicrob. Agents Chemother.* 51, 2351-8.
- (44) Blazer, L. L., and Neubig, R. R. (2009) Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. *Neuropsychopharmacology* 34, 126-41.
- (45) Hibert, M., Trumpp-Kallmeyer, S., Bruinvels, A., and Hoflack, J. (1991) Three-dimensional models of neurotransmitter G-binding protein-coupled receptors. *Mol. Pharmacol.* 40, 8-15.
- (46) Unger, V. M., Hargrave, P. A., Baldwin, J. M., and Schertler, G. F. (1997) Arrangement of rhodopsin transmembrane alpha-helices. *Nature* 389, 203-6.
- (47) Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M., and Miyano, M. (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739-45.
- (48) Hanson, M. A., and Stevens, R. C. (2009) Discovery of new GPCR biology: one receptor structure at a time. *Structure* 17, 8-14.
- (49) Robertson, N., Jazayeri, A., Errey, J., Baig, A., Hurrell, E., Zhukov, A., Langmead, C. J., Weir, M., and Marshall, F. H. (2011) The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* 60, 36-44.
- (50) Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I., Kobilka, B. K., and Stevens, R. C. (2007) High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* 318, 1258-1265.
- (51) Rosenbaum, D. M., Cherezov, V., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Yao, X.-J., Weis, W. I., Stevens, R. C., and Kobilka, B. K. (2007) GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. *Science* 318, 1266-1273.
- (52) Warne, T., Serrano-Vega, M. J., Baker, J. G., Moukhametzianov, R., Edwards, P. C.,

- Henderson, R., Leslie, A. G. W., Tate, C. G., and Schertler, G. F. X. (2008) Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* 454, 486-491.
- (53) Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G., and Tate, C. G. (2011) Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. *Nature* 474, 521-5.
- (54) Xu, F., Wu, H., Katritch, V., Han, G. W., Jacobson, K. A., Gao, Z. G., Cherezov, V., and Stevens, R. C. (2011) Structure of an Agonist-Bound Human A2A Adenosine Receptor. *Science*.
- (55) Congreve, M., Andrews, S. P., Dore, A. S., Hollenstein, K., Hurrell, E., Langmead, C. J., Mason, J. S., Ng, I. W., Tehan, B., Zhukov, A., Weir, M., and Marshall, F. H. (2012) Discovery of 1,2,4-triazine derivatives as adenosine A(2A) antagonists using structure based drug design. *J. Med. Chem.* 55, 1898-903.
- (56) Jaakola, V.-P., Griffith, M. T., Hanson, M. A., Cherezov, V., Chien, E. Y. T., Lane, J. R., Ijzerman, A. P., and Stevens, R. C. (2008) The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* 322, 1211-1217.
- (57) Dore, A. S., Robertson, N., Errey, J. C., Ng, I., Hollenstein, K., Tehan, B., Hurrell, E., Bennett, K., Congreve, M., Magnani, F., Tate, C. G., Weir, M., and Marshall, F. H. (2011) Structure of the adenosine A(2A) receptor in complex with ZM241385 and the xanthines XAC and caffeine. *Structure* 19, 1283-93.
- (58) Warne, T., Moukhametzianov, R., Baker, J. G., Nehme, R., Edwards, P. C., Leslie, A. G., Schertler, G. F., and Tate, C. G. (2011) The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. *Nature* 469, 241-244.
- (59) Moukhametzianov, R., Warne, T., Edwards, P. C., Serrano-Vega, M. J., Leslie, A. G., Tate, C. G., and Schertler, G. F. (2011) Two distinct conformations of helix 6 observed in antagonist-bound structures of a beta1-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* 108, 8228-32.
- (60) Warne, T., Edwards, P. C., Leslie, A. G., and Tate, C. G. (2012) Crystal structures of a stabilized beta1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* 20, 841-9.
- (61) Rasmussen, S. G., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K., and Kobilka, B. K. (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477, 549-55.
- (62) Rosenbaum, D. M., Zhang, C., Lyons, J. A., Holl, R., Aragao, D., Arlow, D. H., Rasmussen, S. G., Choi, H. J., Devree, B. T., Sunahara, R. K., Chae, P. S., Gellman, S. H., Dror, R. O., Shaw, D. E., Weis, W. I., Caffrey, M., Gmeiner, P., and Kobilka, B. K. (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 469, 236-240.
- (63) Wacker, D., Fenalti, G., Brown, M. A., Katritch, V., Abagyan, R., Cherezov, V., and Stevens, R. C. (2010) Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* 132, 11443-5.
- (64) Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V. P., Chien, E. Y., Velasquez, J., Kuhn, P., and Stevens, R. C. (2008) A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Structure* 16, 897-905.
- (65) Wu, B., Chien, E. Y., Mol, C. D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F. C., Hamel, D. J., Kuhn, P., Handel, T. M., Cherezov, V., and Stevens, R. C. (2010) Structures of the CXCR4 chemokine GPCR with small-

- molecule and cyclic peptide antagonists. *Science* 330, 1066-1071.
- (66) Chien, E. Y. T., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., and Stevens, R. C. (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330, 1091-1095.
- (67) Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katritch, V., Abagyan, R., Cherezov, V., Liu, W., Han, G. W., Kobayashi, T., Stevens, R. C., and Iwata, S. (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475, 65-70.
- (68) Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shiroishi, M., Zhang, C., Weis, W. I., Okada, T., Kobilka, B. K., Haga, T., and Kobayashi, T. (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482, 547-51.
- (69) Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., Dror, R. O., Shaw, D. E., Weis, W. I., Wess, J., and Kobilka, B. K. (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552-6.
- (70) Granier, S., Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Weis, W. I., and Kobilka, B. K. (2012) Structure of the delta-opioid receptor bound to naltrindole. *Nature* 485, 400-4.
- (71) Manglik, A., Kruse, A. C., Kobilka, T. S., Thian, F. S., Mathiesen, J. M., Sunahara, R. K., Pardo, L., Weis, W. I., Kobilka, B. K., and Granier, S. (2012) Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* 485, 321-6.
- (72) Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. A., Huang, X. P., Carroll, F. I., Mascarella, S. W., Westkaemper, R. B., Mosier, P. D., Roth, B. L., Cherezov, V., and Stevens, R. C. (2012) Structure of the human kappa-opioid receptor in complex with JDTC. *Nature* 485, 327-32.
- (73) Thompson, A. A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., Huang, X.-P., Trapella, C., Guerrini, R., Calo, G., Roth, B. L., Cherezov, V., and Stevens, R. C. (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395-399.
- (74) Hanson, M. A., Roth, C. B., Jo, E., Griffith, M. T., Scott, F. L., Reinhart, G., Desale, H., Clemons, B., Cahalan, S. M., Schuerer, S. C., Sanna, M. G., Han, G. W., Kuhn, P., Rosen, H., and Stevens, R. C. (2012) Crystal structure of a lipid G protein-coupled receptor. *Science* 335, 851-5.
- (75) Choe, H. W., Kim, Y. J., Park, J. H., Morizumi, T., Pai, E. F., Krauss, N., Hofmann, K. P., Scheerer, P., and Ernst, O. P. (2011) Crystal structure of metarhodopsin II. *Nature* 471, 651-5.
- (76) Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W., and Ernst, O. P. (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454, 183-187.
- (77) Okada, T., Sugihara, M., Bondar, A. N., Elstner, M., Entel, P., and Buss, V. (2004) The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* 342, 571-83.
- (78) Makino, C. L., Riley, C. K., Looney, J., Crouch, R. K., and Okada, T. (2010) Binding of more than one retinoid to visual opsins. *Biophys. J.* 99, 2366-73.
- (79) Murakami, M., and Kouyama, T. (2011) Crystallographic analysis of the primary photochemical reaction of squid rhodopsin. *J. Mol. Biol.* 413, 615-27.
- (80) Murakami, M., and Kouyama, T. (2008) Crystal structure of squid rhodopsin. *Nature* 453, 363-7.
- (81) Westbrook, J., Feng, Z. K., Chen, L., Yang, H. W., and Berman, H. M. (2003) The

- Protein Data Bank and structural genomics. *Nucleic Acids Res.* 31, 489-491.
- (82) Katritch, V., Cherezov, V., and Stevens, R. C. (2011) Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* 33, 17-27.
- (83) Congreve, M., Langmead, C. J., Mason, J. S., and Marshall, F. H. (2011) Progress in structure based drug design for G protein-coupled receptors. *J. Med. Chem.* 54, 4283-311.
- (84) Brothers, S. P., and Wahlestedt, C. (2010) Therapeutic potential of neuropeptide Y (NPY) receptor ligands. *EMBO Mol Med* 2, 429-39.
- (85) Conn, P. M., and Ulloa-Aguirre, A. (2011) Pharmacological chaperones for misfolded gonadotropin-releasing hormone receptors. *Adv. Pharmacol.* 62, 109-41.
- (86) Zingg, H. H., and Laporte, S. A. (2003) The oxytocin receptor. *Trends Endocrinol Metab* 14, 222-7.
- (87) Jacobson, K. A., Deflorian, F., Mishra, S., and Costanzi, S. (2011) Pharmacology of the platelet purinergic receptors. *Purinergic Signal* 7, 305-24.
- (88) Ballesteros, J. A., and Weinstein, H. (1995) Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors, in *Methods Neurosci.* pp 366-428, Academic Press, San Diego.
- (89) Tieleman, D. P., Shrivastava, I. H., Ulmschneider, M. R., and Sansom, M. S. (2001) Proline-induced hinges in transmembrane helices: possible roles in ion channel gating. *Proteins* 44, 63-72.
- (90) Ballesteros, J. A., Jensen, A. D., Liapakis, G., Rasmussen, S. G., Shi, L., Gether, U., and Javitch, J. A. (2001) Activation of the beta 2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* 276, 29171-29177.
- (91) Vogel, R., Mahalingam, M., Lüdeke, S., Huber, T., Siebert, F., and Sakmar, T. P. (2008) Functional role of the "ionic lock"--an interhelical hydrogen-bond network in family A heptahelical receptors. *J. Mol. Biol.* 380, 648-655.
- (92) Audet, M., and Bouvier, M. (2008) Insights into signaling from the beta2-adrenergic receptor structure. *Nat. Chem. Biol.* 4, 397-403.
- (93) Ruprecht, J. J., Mielke, T., Vogel, R., Villa, C., and Schertler, G. F. (2004) Electron crystallography reveals the structure of metarhodopsin I. *EMBO J.* 23, 3609-20.
- (94) Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M., and Schwartz, T. W. (2009) Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol. Sci.* 30, 249-259.
- (95) Shi, L., Liapakis, G., Xu, R., Guarnieri, F., Ballesteros, J. A., and Javitch, J. A. (2002) Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* 277, 40989-40996.
- (96) Kobilka, B. K., and Deupi, X. (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* 28, 397-406.
- (97) Pardo, L., Deupi, X., Dölker, N., López-Rodríguez, M. L., and Campillo, M. (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *ChemBioChem* 8, 19-24.
- (98) Angel, T. E., Chance, M. R., and Palczewski, K. (2009) Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8555-8560.
- (99) Angel, T. E., Gupta, S., Jastrzebska, B., Palczewski, K., and Chance, M. R. (2009) Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* 106, 14367-14372.
- (100) Nygaard, R., Valentin-Hansen, L., Mokrosinski, J., Frimurer, T. M., and Schwartz, T.

- W. (2010) Conserved water-mediated hydrogen bond network between TM-I, -II, -VI, and -VII in 7TM receptor activation. *J. Biol. Chem.* 285, 19625-36.
- (101) Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., Han, G. W., Roth, C. B., Heitman, L. H., AP, I. J., Cherezov, V., and Stevens, R. C. (2012) Structural basis for allosteric regulation of GPCRs by sodium ions. *Science* 337, 232-6.
- (102) Deupi, X., and Standfuss, J. (2011) Structural insights into agonist-induced activation of G-protein-coupled receptors. *Curr. Opin. Struct. Biol.* 21, 541-51.
- (103) Trzaskowski, B., Latek, D., Yuan, S., Ghoshdastider, U., Debinski, A., and Filipek, S. (2012) Action of molecular switches in GPCRs--theoretical and experimental studies. *Curr. Med. Chem.* 19, 1090-109.
- (104) Rodriguez, D., Pineiro, A., and Gutierrez-de-Teran, H. (2011) Molecular dynamics simulations reveal insights into key structural elements of adenosine receptors. *Biochemistry* 50, 4194-208.
- (105) Filipek, S., Teller, D. C., Palczewski, K., and Stenkamp, R. (2003) The Crystallographic Model of Rhodopsin and Its Use in Studies of Other G Protein-Coupled Receptors. *Annu. Rev. Biophys. Biomol. Struct.* 32, 375-397.
- (106) Hopkins, A. L., and Groom, C. R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727-730.
- (107) Shoichet, B. K., and Kobilka, B. K. (2012) Structure-based drug screening for G-protein-coupled receptors. *Trends Pharmacol. Sci.* 33, 268-72.
- (108) de Graaf, C., and Rognan, D. (2008) Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor. *J. Med. Chem.* 51, 4978-85.
- (109) Jaakola, V. P., Lane, J. R., Lin, J. Y., Katritch, V., Ijzerman, A. P., and Stevens, R. C. (2010) Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: identification and characterization of essential amino acid residues. *J. Biol. Chem.* 285, 13032-13044.
- (110) Katritch, V., Jaakola, V. P., Lane, J. R., Lin, J., Ijzerman, A. P., Yeager, M., Kufareva, I., Stevens, R. C., and Abagyan, R. (2010) Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. *J. Med. Chem.* 53, 1799-809.
- (111) Kim, J., Wess, J., van Rhee, A. M., Schöneberg, T., and Jacobson, K. A. (1995) Site-directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A2A Adenosine Receptor. *J. Biol. Chem.* 270, 13987-13997.
- (112) Gao, Z. G., Jiang, Q., Jacobson, K. A., and Ijzerman, A. P. (2000) Site-directed mutagenesis studies of human A(2A) adenosine receptors: involvement of Glu(13) and His(278) in ligand binding and sodium modulation. *Biochem. Pharmacol.* 60, 661-668.
- (113) Maggio, R., Vogel, Z., and Wess, J. (1993) Coexpression studies with mutant muscarinic/adrenergic receptors provide evidence for intermolecular "cross-talk" between G-protein-linked receptors. *Proc. Natl. Acad. Sci. U. S. A.* 90, 3103-7.
- (114) Fanelli, F., and De Benedetti, P. G. (2011) Update 1 of: computational modeling approaches to structure-function analysis of G protein-coupled receptors. *Chem Rev* 111, PR438-535.
- (115) George, S. R., O'Dowd, B. F., and Lee, S. R. (2002) G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat. Rev. Drug Discovery* 1, 808-820.
- (116) Gonzalez-Maeso, J., Ang, R. L., Yuen, T., Chan, P., Weisstaub, N. V., Lopez-Gimenez, J. F., Zhou, M., Okawa, Y., Callado, L. F., Milligan, G., Gingrich, J. A., Filizola, M., Meana, J. J., and Sealfon, S. C. (2008) Identification of a serotonin/glutamate receptor complex implicated in psychosis. *Nature* 452, 93-7.

- (117) Fuxe, K., Marcellino, D., Genedani, S., and Agnati, L. (2007) Adenosine A(2A) receptors, dopamine D(2) receptors and their interactions in Parkinson's disease. *Mov Disord* 22, 1990-2017.
- (118) Selent, J., and Kaczor, A. A. (2011) Oligomerization of G protein-coupled receptors: computational methods. *Curr. Med. Chem.* 18, 4588-605.
- (119) Liang, Y., Fotiadis, D., Filipek, S., Saperstein, D. A., Palczewski, K., and Engel, A. (2003) Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J. Biol. Chem.* 278, 21655-62.
- (120) Fotiadis, D., Liang, Y., Filipek, S., Saperstein, D. A., Engel, A., and Palczewski, K. (2003) Atomic-force microscopy: Rhodopsin dimers in native disc membranes. *Nature* 421, 127-8.
- (121) Filizola, M. (2010) Increasingly accurate dynamic molecular models of G-protein coupled receptor oligomers: Panacea or Pandora's box for novel drug discovery? *Life Sci.* 86, 590-7.
- (122) Salom, D., Lodowski, D. T., Stenkamp, R. E., Le Trong, I., Golczak, M., Jastrzebska, B., Harris, T., Ballesteros, J. A., and Palczewski, K. (2006) Crystal structure of a photoactivated deprotonated intermediate of rhodopsin. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16123-8.
- (123) Wang, J., and Norcross, M. (2008) Dimerization of chemokine receptors in living cells: key to receptor function and novel targets for therapy. *Drug Discov. Today* 13, 625-32.
- (124) Babcock, G. J., Farzan, M., and Sodroski, J. (2003) Ligand-independent dimerization of CXCR4, a principal HIV-1 coreceptor. *J. Biol. Chem.* 278, 3378-3385.
- (125) Wang, J., He, L., Combs, C. A., Roderiquez, G., and Norcross, M. A. (2006) Dimerization of CXCR4 in living malignant cells: control of cell migration by a synthetic peptide that reduces homologous CXCR4 interactions. *Mol. Cancer Ther.* 5, 2474-83.
- (126) Kaczor, A. A., and Selent, J. (2011) Oligomerization of G protein-coupled receptors: biochemical and biophysical methods. *Curr. Med. Chem.* 18, 4606-34.
- (127) de Graaf, C., and Rognan, D. (2009) Customizing G Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* 15, 4026-48.
- (128) Vaidehi, N., Floriano, W. B., Trabaino, R., Hall, S. E., Freddolino, P., Choi, E. J., Zamanakos, G., and Goddard, W. A. (2002) Prediction of structure and function of G protein-coupled receptors. *PNAS* 99, 12622-12627.
- (129) Shacham, S., Topf, M., Avisar, N., Glaser, F., Marantz, Y., Bar-Haim, S., Noiman, S., Naor, Z., and Becker, O. M. (2001) Modeling the 3D structure of GPCRs from sequence. *Med. Res. Rev.* 21, 472-83.
- (130) Carrieri, A., Centeno, N. B., Rodrigo, J., Sanz, F., and Carotti, A. (2001) Theoretical evidence of a salt bridge disruption as the initiating process for the α 1D-adrenergic receptor activation: a molecular dynamics and docking study. *Proteins* 43, 382-394.
- (131) Goddard, W. A., 3rd, and Abrol, R. (2007) 3-Dimensional structures of G protein-coupled receptors and binding sites of agonists and antagonists. *J. Nutr.* 137, 1528S-1538S; discussion 1548S.
- (132) Michino, M., Chen, J., Stevens, R. C., and Brooks, C. L., 3rd. (2010) FoldGPCR: structure prediction protocol for the transmembrane domain of G protein-coupled receptors from class A. *Proteins* 78, 2189-2201.
- (133) Spijker, P., Vaidehi, N., Freddolino, P. L., Hilbers, P. A., and Goddard, W. A., 3rd. (2006) Dynamic behavior of fully solvated beta2-adrenergic receptor, embedded in the membrane with bound agonist or antagonist. *Proc. Natl. Acad. Sci. U. S. A.* 103, 4882-7.

-
- (134) Zhang, Y., Devries, M. E., and Skolnick, J. (2006) Structure modeling of all identified G protein-coupled receptors in the human genome. *PLoS Comput. Biol.* 2, e13.
- (135) Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Sali, A. (2000) Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* 29, 291-325.
- (136) Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* 49, 2987-98.
- (137) Michino, M., Abola, E., Brooks, C. L., Dixon, J. S., Moulton, J., and Stevens, R. C. (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* 8, 455-463.
- (138) Kufareva, I., Rueda, M., Katritch, V., Stevens, R. C., and Abagyan, R. (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* 19, 1108-26.
- (139) Costanzi, S. (2012) Homology modeling of class a G protein-coupled receptors. *Methods Mol. Biol.* 857, 259-79.
- (140) de Graaf, C., Foata, N., Engkvist, O., and Rognan, D. (2008) Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins* 71, 599-620.
- (141) Cavasotto, C. N., Orry, A. J., Murgolo, N. J., Czarniecki, M. F., Kocsi, S. A., Hawes, B. E., O'Neill, K. A., Hine, H., Burton, M. S., Voigt, J. H., Abagyan, R. A., Bayne, M. L., and Monsma, F. J., Jr. (2008) Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* 51, 581-8.
- (142) Phatak, S. S., Gatica, E. A., and Cavasotto, C. N. (2010) Ligand-steered modeling and docking: A benchmarking study in class A G-protein-coupled receptors. *J. Chem. Inf. Model.* 50, 2119-28.
- (143) Katritch, V., Rueda, M., Lam, P. C., Yeager, M., and Abagyan, R. (2010) GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. *Proteins* 78, 197-211.
- (144) Rodríguez, D., Bello, X., and Gutiérrez-de-Terán, H. (2012) Molecular Modelling of G Protein-Coupled Receptors Through the Web. *Molecular Informatics* 31, 334-341.
- (145) Heifetz, A., Morris, G. B., Biggin, P. C., Barker, O., Fryatt, T., Bentley, J., Hallett, D., Manikowski, D., Pal, S., Reifegerste, R., Slack, M., and Law, R. (2012) Study of human Orexin-1 and -2 G-protein-coupled receptors with novel and published antagonists by modeling, molecular dynamics simulations, and site-directed mutagenesis. *Biochemistry* 51, 3178-97.
- (146) Kimura, S. R., Tebben, A. J., and Langley, D. R. (2008) Expanding GPCR homology model binding sites via a balloon potential: A molecular dynamics refinement approach. *Proteins* 71, 1919-29.
- (147) Mobarec, J. C., Sanchez, R., and Filizola, M. (2009) Modern homology modeling of G-protein coupled receptors: which structural template to use? *J. Med. Chem.* 52, 5207-5216.
- (148) Worth, C. L., Kreuchwig, A., Kleinau, G., and Krause, G. (2011) GPCR-SSFE: a comprehensive database of G-protein-coupled receptor template predictions and homology models. *BMC Bioinformatics* 12, 185.
- (149) Worth, C. L., Kleinau, G., and Krause, G. (2009) Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One* 4, e7011.
- (150) Rost, B. (1999) Twilight zone of protein sequence alignments. *Protein Eng.* 12, 85-

- 94.-85-94.
- (151) Carlsson, J., Coleman, R. G., Setola, V., Irwin, J. J., Fan, H., Schlessinger, A., Sali, A., Roth, B. L., and Shoichet, B. K. (2011) Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nat. Chem. Biol.* 7, 769-78.
- (152) Langmead, C. J., Andrews, S. P., Congreve, M., Errey, J. C., Hurrell, E., Marshall, F. H., Mason, J. S., Richardson, C. M., Robertson, N., Zhukov, A., and Weir, M. (2011) Identification of novel adenosine A(2A) receptor antagonists by virtual screening. *J. Med. Chem.* 55, 1904-9.
- (153) Tang, H., Wang, X. S., Hsieh, J. H., and Tropsha, A. (2012) Do crystal structures obviate the need for theoretical models of GPCRs for structure-based virtual screening? *Proteins* 80, 1503-21.
- (154) Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947-8.
- (155) Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.
- (156) Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.* 9, 1753-1773.
- (157) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283-291.
- (158) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375-383.
- (159) Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665-667.
- (160) Macromodel, version 9.7, New York, NY, 2009.
- (161) Song, Y., Mao, J., and Gunner, M. R. (2009) MCCE2: improving protein pKa calculations with extensive side chain rotamer sampling. *J. Comput. Chem.* 30, 2231-2247.
- (162) Pieper, U., Webb, B. M., Barkan, D. T., Schneidman-Duhovny, D., Schlessinger, A., Braberg, H., Yang, Z., Meng, E. C., Pettersen, E. F., Huang, C. C., Datta, R. S., Sampathkumar, P., Madhusudhan, M. S., Sjolander, K., Ferrin, T. E., Burley, S. K., and Sali, A. (2011) ModBase, a database of annotated comparative protein structure models, and associated resources. *Nucleic Acids Res.* 39, D465-74.
- (163) Kiefer, F., Arnold, K., Kunzli, M., Bordoli, L., and Schwede, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res.* 37, D387-92.
- (164) Zhang, J., and Zhang, Y. (2010) GPCRDR: G protein-coupled receptor spatial restraint database for 3D structure modeling and function annotation. *Bioinformatics* 26, 3004-5.
- (165) Zhou, H., and Skolnick, J. (2012) FINDSITE(X): A Structure-Based, Small Molecule Virtual Screening Approach with Application to All Identified Human GPCRs. *Mol. Pharm.*
- (166) Kelm, S., Shi, J., and Deane, C. M. (2010) MEDELLER: homology-based coordinate generation for membrane proteins. *Bioinformatics* 26, 2833-40.
- (167) Vroiling, B., Sanders, M., Baakman, C., Borrmann, A., Verhoeven, S., Klomp, J., Oliveira, L., de Vlieg, J., and Vriend, G. (2011) GPCRDB: information system for G

- protein-coupled receptors. *Nucleic Acids Res.* 39, D309-19.
- (168) Yuriev, E., Agostino, M., and Ramsland, P. A. (2011) Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* 24, 149-64.
- (169) Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2006) Protein-ligand docking: current status and future challenges. *Proteins* 65, 15-26.
- (170) Cheng, T., Li, Q., Zhou, Z., Wang, Y., and Bryant, S. H. (2012) Structure-based virtual screening for drug discovery: a problem-centric review. *Aaps J* 14, 133-41.
- (171) Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein-ligand docking using GOLD. *Proteins-Structure Function and Genetics* 52, 609-623.
- (172) Baxter, C. A., Murray, C. W., Clark, D. E., Westhead, D. R., and Eldridge, M. D. (1998) Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* 33, 367-82.
- (173) Sabio, M., Jones, K., and Topiol, S. (2008) Use of the X-ray structure of the beta2-adrenergic receptor for drug discovery. Part 2: Identification of active compounds. *Bioorg. Med. Chem. Lett.* 18, 5391-5.
- (174) Kolb, P., Rosenbaum, D. M., Irwin, J. J., Fung, J. J., Kobilka, B. K., and Shoichet, B. K. (2009) Structure-based discovery of beta2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* 106, 6843-6848.
- (175) Carlsson, J., Yoo, L., Gao, Z.-G., Irwin, J. J., Shoichet, B. K., and Jacobson, K. A. (2010) Structure-based discovery of A2A adenosine receptor ligands. *J. Med. Chem.* 53, 3748-3755.
- (176) Mysinger, M. M., Weiss, D. R., Ziarek, J. J., Gravel, S., Doak, A. K., Karpiak, J., Heveker, N., Shoichet, B. K., and Volkman, B. F. (2012) Structure-based ligand discovery for the protein-protein interface of chemokine receptor CXCR4. *Proc. Natl. Acad. Sci. U. S. A.* 109, 5517-22.
- (177) de Graaf, C., Kooistra, A. J., Vischer, H. F., Katritch, V., Kuijper, M., Shiroishi, M., Iwata, S., Shimamura, T., Stevens, R. C., de Esch, I. J., and Leurs, R. (2011) Crystal structure-based virtual screening for fragment-like ligands of the human histamine H(1) receptor. *J. Med. Chem.* 54, 8195-206.
- (178) Costanzi, S., Tikhonova, I. G., Harden, T. K., and Jacobson, K. A. (2009) Ligand and structure-based methodologies for the prediction of the activity of G protein-coupled receptor ligands. *J. Comput Aided Mol Des* 23, 747-54.
- (179) Bender, A., and Glen, R. C. (2004) Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* 2, 3204-18.
- (180) Eckert, H., and Bajorath, J. (2007) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today* 12, 225-33.
- (181) Rush, T. S., 3rd, Grant, J. A., Mosyak, L., and Nicholls, A. (2005) A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. *J. Med. Chem.* 48, 1489-95.
- (182) Hawkins, P. C., Skillman, A. G., Warren, G. L., Ellingson, B. A., and Stahl, M. T. (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* 50, 572-84.
- (183) Hawkins, P. C. D., Skillman, A. G., and Nicholls, A. (2007) Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* 50, 74-82.
- (184) VIDA, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, 2008.
- (185) Esposito, E. X., Hopfinger, A. J., and Madura, J. D. (2004) Methods for applying the quantitative structure-activity relationship paradigm. *Methods Mol. Biol.* 275, 131-

- 214.
- (186) Goodford, P. J. (1985) A computational procedure for determining energetically favorable binding sites of biological important macromolecules. *Journal of Medicinal Chemistry*. 28, 849-857.
- (187) Moro, S., Braiuca, P., Deflorian, F., Ferrari, C., Pastorin, G., Cacciari, B., Baraldi, P. G., Varani, K., Borea, P. A., and Spalluto, G. (2005) Combined target-based and ligand-based drug design approach as a tool to define a novel 3D-pharmacophore model of human A3 adenosine receptor antagonists: pyrazolo[4,3-e]1,2,4-triazolo[1,5-c]pyrimidine derivatives as a key study. *J. Med. Chem.* 48, 152-62.
- (188) Duran, A., Martinez, G. C., and Pastor, M. (2008) Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in Molecular Interaction Fields. *J. Chem. Inf. Model.* 48, 1813-23.
- (189) Bender, A., Mussa, H. Y., Glen, R. C., and Reiling, S. (2004) Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance. *J Chem Inf Comput Sci* 44, 1708-18.
- (190) Bender, A., Mussa, H. Y., Glen, R. C., and Reiling, S. (2004) Molecular similarity searching using atom environments, information-based feature selection, and a naive Bayesian classifier. *J. Chem. Inf. Comput. Sci.* 44, 170-8.
- (191) Razzak, A. M., and Glen, R. C. (1992) Applications of rule-induction in the derivation of quantitative structure-activity relationships. *J Comput Aided Mol Des* 6, 349-83.
- (192) Bender, A., Mussa, H. Y., and Glen, R. C. (2005) Screening for dihydrofolate reductase inhibitors using MOLPRINT 2D, a fast fragment-based method employing the naive Bayesian classifier: limitations of the descriptor and the importance of balanced chemistry in training and test sets. *J. Biomol. Screen.* 10, 658-66.
- (193) Koutsoukas, A., Simms, B., Kirchmair, J., Bond, P. J., Whitmore, A. V., Zimmer, S., Young, M. P., Jenkins, J. L., Glick, M., Glen, R. C., and Bender, A. (2012) From in silico target prediction to multi-target drug design: current databases, methods and applications. *J Proteomics* 74, 2554-74.
- (194) Guvench, O., and MacKerell, A. D., Jr. (2008) Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* 443, 63-88.
- (195) Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225-11236.
- (196) Berman, H. M. (2008) The Protein Data Bank: a historical perspective. *Acta Crystallogr A* 64, 88-95.
- (197) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089-10089.
- (198) van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., and Berendsen, H. J. C. (2010) *Gromacs User Manual version 4.5.4*.
- (199) Lindahl, E., and Sansom, M. S. (2008) Membrane proteins: molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 18, 425-31.
- (200) Park, P. S., Lodowski, D. T., and Palczewski, K. (2008) Activation of G protein-coupled receptors: beyond two-state models and tertiary conformational changes. *Annu. Rev. Pharmacol. Toxicol.* 48, 107-41.
- (201) Bokoch, M. P., Zou, Y., Rasmussen, S. G. F., Liu, C. W., Nygaard, R., Rosenbaum, D. M., Fung, J. J., Choi, H.-J., Thian, F. S., Kobilka, T. S., Puglisi, J. D., Weis, W. I., Pardo, L., Prosser, R. S., Mueller, L., and Kobilka, B. K. (2010) Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor. *Nature* 463, 108-112.

- (202) Liu, J. J., Horst, R., Katritch, V., Stevens, R. C., and Wuthrich, K. (2012) Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR. *Science* 335, 1106-10.
- (203) Hubbell, W. L., Altenbach, C., Hubbell, C. M., and Khorana, H. G. (2003) Rhodopsin structure, dynamics, and activation: a perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking. *Adv. Protein Chem.* 63, 243-90.
- (204) Kandt, C., Ash, W. L., and Tieleman, D. P. (2007) Setting up and running molecular dynamics simulations of membrane proteins. *Methods (San Diego, Calif.)* 41, 475-488.
- (205) Biggin, P. C., and Bond, P. J. (2008) Molecular dynamics simulations of membrane proteins. *Methods Mol. Biol.* 443, 147-60.
- (206) Grossfield, A. (2011) Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations. *Biochim. Biophys. Acta* 1808, 1868-78.
- (207) Johnston, J. M., and Filizola, M. (2011) Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. *Curr. Opin. Struct. Biol.* 21, 552-8.
- (208) Bruno, A., and Costantino, G. (2012) Molecular Dynamics Simulations of G Protein-Coupled Receptors. *Molecular Informatics* 31, 222-230.
- (209) Lomize, M. A., Lomize, A. L., Pogozheva, I. D., and Mosberg, H. I. (2006) OPM: orientations of proteins in membranes database. *Bioinformatics* 22, 623-5.
- (210) Díaz, L., Bujons, J., Delgado, A., Gutiérrez-de-Terán, H., and Aqvist, J. (2011) Computational prediction of structure-activity relationships for the binding of aminocyclitols to beta-glucocerebrosidase. *J. Chem. Inf. Model.* 51, 601-11.
- (211) Tieleman, D. P., Maccallum, J. L., Ash, W. L., Kandt, C., Xu, Z., and Monticelli, L. (2006) Membrane protein simulations with a united-atom lipid and all-atom protein model: lipid-protein interactions, side chain transfer free energies and model proteins. *J Phys Condens Matter* 18, S1221-34.
- (212) Berger, O., Edholm, O., and Jähnig, F. (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72, 2002-2013.
- (213) Chakrabarti, N., Neale, C., Payandeh, J., Pai, E. F., and Pomès, R. (2010) An iris-like mechanism of pore dilation in the CorA magnesium transport system. *Biophys. J.* 98, 784-792.
- (214) Anézo, C., de Vries, A. H., Höltje, H.-D., Tieleman, D. P., and Marrink, S.-J. (2003) Methodological Issues in Lipid Bilayer Simulations. *J. Phys. Chem. B* 107, 9424-9433.
- (215) Selent, J., López, L., Sanz, F., and Pastor, M. (2008) Multi-receptor binding profile of clozapine and olanzapine: a structural study based on the new beta2 adrenergic receptor template. *ChemMedChem* 3, 1194-1198.
- (216) Katritch, V., Kufareva, I., and Abagyan, R. (2011) Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharmacology* 60, 108-15.
- (217) Gutiérrez de Terán, H. Ph.D. thesis, Universitat Pompeu Fabra, Barcelona, 2004. <http://hdl.handle.net/10803/7075>
- (218) Pentacle 1.1, Molecular Discovery, Ltd, Perugia, Italy, 2009.
- (219) van der Horst, E., van der Pijl, R., Mulder-Krieger, T., Bender, A., and Ijzerman, A. P. (2011) Substructure-based virtual screening for adenosine A2A receptor ligands. *ChemMedChem* 6, 2302-11.

- (220) Sanders, M. P., Roumen, L., van der Horst, E., Lane, J. R., Vischer, H. F., van Offenbeek, J., de Vries, H., Verhoeven, S., Chow, K. Y., Verkaar, F., Beukers, M. W., McGuire, R., Leurs, R., Ijzerman, A. P., de Vlieg, J., de Esch, I. J., Zaman, G. J., Klomp, J. P., Bender, A., and de Graaf, C. (2012) A Prospective Cross-Screening Study on G-Protein-Coupled Receptors: Lessons Learned in Virtual Compound Library Design. *J. Med. Chem.*
- (221) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4, 435-447.
- (222) Lyman, E., Higgs, C., Kim, B., Lupyán, D., Shelley, J. C., Farid, R., and Voth, G. A. (2009) A role for a specific cholesterol interaction in stabilizing the Apo configuration of the human A(2A) adenosine receptor. *Structure* 17, 1660-1668.
- (223) Huber, T., Menon, S., and Sakmar, T. P. (2008) Structural basis for ligand binding and specificity in adrenergic receptors: implications for GPCR-targeted drug discovery. *Biochemistry* 47, 11013-11023.
- (224) Dror, R. O., Arlow, D. H., Borhani, D. W., Jensen, M. Ø., Piana, S., and Shaw, D. E. (2009) Identification of two distinct inactive conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U. S. A.* 106, 4689-4694.
- (225) Vanni, S., Neri, M., Tavernelli, I., and Rothlisberger, U. (2009) Observation of "ionic lock" formation in molecular dynamics simulations of wild-type beta(1) and beta(2) adrenergic receptors. *Biochemistry* 48, 4789-4797.
- (226) Vanni, S., Neri, M., Tavernelli, I., and Rothlisberger, U. (2010) A conserved protonation-induced switch can trigger "ionic-lock" formation in adrenergic receptors. *J. Mol. Biol.* 397, 1339-1349.
- (227) Romo, T. D., Grossfield, A., and Pitman, M. C. (2010) Concerted interconversion between ionic lock substates of the beta(2) adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* 98, 76-84.
- (228) Dror, R. O., Arlow, D. H., Maragakis, P., Mildorf, T. J., Pan, A. C., Xu, H., Borhani, D. W., and Shaw, D. E. (2011) Activation mechanism of the beta2-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* 108, 18684-9.
- (229) Selent, J., Sanz, F., Pastor, M., and De Fabritiis, G. (2010) Induced effects of sodium ions on dopaminergic G-protein coupled receptors. *PLoS Comput. Biol.* 6, e1000884.
- (230) Gonzalez-Maeso, J. (2011) GPCR oligomers in pharmacology and signaling. *Mol Brain* 4, 20.
- (231) Zhang, Y., and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* 33, 2302-9.
- (232) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- (233) Fallmar, H., Kerberg, H., Gutierrez-de-Teran, H., Lundell, I., Mohell, N., and Larhammar, D. (2011) Identification of positions in the human neuropeptide Y/peptide YY receptor Y2 that contribute to pharmacological differences between receptor subtypes. *Neuropeptides* 45, 293-300.
- (234) Shen, M. Y., and Sali, A. (2006) Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 15, 2507-24.
- (235) Wada, M., Kanamori, E., Nakamura, H., and Fukunishi, Y. (2011) Selection of in silico drug screening results for G-protein-coupled receptors by using universal active probes. *J. Chem. Inf. Model.* 51, 2398-407.

7. PUBLICATIONS AND SUPPORTING INFORMATION

Paper I

Gutiérrez-de-Terán H., Correia C., **Rodríguez D.**, Carvalho M.A., Brea J., Cadavid M.I., Loza M.I., Proença M.F. and Areias F.

Identification of novel scaffolds from an original chemical library as potential antipsychotics

QSAR & Comb. Sci., **2009** 28(8):856-60

Identification of Novel Scaffolds from an Original Chemical Library as Potential Antipsychotics

Hugo Gutiérrez-de-Terán,^a Carla Correia,^c David Rodríguez,^a Maria Alice Carvalho,^c Jose Brea,^b María Isabel Cadavid,^b María Isabel Loza,^b Maria Fernanda Proença,^c Filipe Areias^{b,c,*}

^a Fundación Pública Galega de Medicina Xenómica. Hospital Clínico Universitario de Santiago de Compostela, Edificio Consultas Planta – 2, 15706 Santiago de Compostela, Spain

^b Departamento de Farmacología, Instituto de Farmacia Industrial. Faculdade de Farmacia, Universidade de Santiago de Compostela, Campus sur, s/n 15706 Santiago de Compostela, Spain

^c Centro de Química, Universidade do Minho. Campus de Gualtar 4710-057 Braga, Portugal

*E-mail: filipemiguel.ferreira@usc.es

Keywords: Antipsychotic, Virtual screening, Purine, Serotonin receptors, 5-HT_{2A}, Drug design

Presented at the EuroQSAR Conference

Received: December 15, 2008; Accepted: March 20, 2009

DOI: 10.1002/qsar.200860198


Abstract

The ligand-based virtual screening of an original chemical library, using atypical antipsychotics as query compounds led to the identification of a novel scaffold with inhibitory activity at the 5-HT_{2A} serotonin receptor. The hit compounds were confirmed by pharmacological evaluation at the 5-HT_{2A} receptor and complemented by the selection of other representatives of the same chemical family within our chemical library. A promising scaffold of 6-(pyperazin-1-yl) purine was identified, and the binding mode is illustrated with an automated docking exploration on a homology built model of the 5-HT_{2A} receptor. The present results constitute an excellent starting point for the discovery of new chemical entities with antipsychotic activity.

Schizophrenia is a severe disorder that affects around 24 million people worldwide, typically beginning in late adolescence or early adulthood. It constitutes one of the major psychiatric disorders in the world, and it can impair functioning through the loss of the acquired capability of earning one's own livelihood or through the disruption of studies [1]. Up to date, pharmacological therapy with antipsychotics constitutes the most effective way to maintain most of the symptoms under control. Nowadays, clozapine, discovered nearly 50 years ago, remains as the gold standard antipsychotic, being the only licensed drug indicated for treatment-resistant schizophrenia. However, its use is restricted to these cases mainly due to the risk of agranulocytosis, its major adverse effect [2]. Based on its high affinity at the serotonin 5-HT_{2A} receptors, modulating its intermediate affinity over dopamine D₂ receptor, a putative mechanism of action of the so-called atypical antipsychotic drugs has been primary established by the Meltzer index, which is the relationship between the affinities at the aforementioned receptors [3]. However, recent multireceptorial profiling of clozapine has shown its affinity at several aminergic G protein-coupled receptors (GPCRs) [4], opening a new scenario where the beneficial effects of

atypical antipsychotics on cognition, negative symptoms, and the low incidence of EPS are mediated by a complex blend of interactions. Therefore, there has been growing interest in the discovery of new compounds that exhibit a similar pharmacological profile as clozapine, but possessing clozapine-unrelated chemical structures.

Three dimensional ligand-based virtual screening (3D-LBVS) represents a good strategy to retrieve original, chemically different compounds from a chemical database, displaying similar pharmacological profile to a given compound with clinical interest [5, 6]. This statement is not in disagreement with the excellent performance of 2D-based LBVS methodologies, as it has been noted by several authors [6, 7]. For the particular case of antipsychotics, however, 3D-LBVS appears as an especially well suited technique attending to the following reasons: i) The target proteins are several aminergic receptors of the GPCR superfamily. Even if the traditional handicap of lacking crystal structures of aminergic GPCRs has very recently been partially overcome [8, 9], we still depend on the generation of

 Supporting information for this article is available on the WWW under www.qcs.wiley-vch.de

homology-derived models to perform receptor based (RB) VS. ii) The multireceptorial profile, responsible for the clinical efficacy of this type of drugs, [10] points out that one should in principle screen against a battery of receptors. Nevertheless, other alternative approaches have been proposed for the design of novel scaffolds for GPCR ligands with polipharmacology, which include hybrid structure based method [11], or a proper combination of physicochemical descriptors that characterize the binders of each considered receptor [12].

The present work is part of a collaborative project dealing with an exhaustive pharmacological characterization of an original chemical library, in order to find hit compounds for further development and lead optimization in different targets. We herein focus on the identification of a new family of compounds with promising affinities at the 5-HT_{2A} receptors, using LBVS. Hit compounds are confirmed by binding assays further and characterized on the basis of a receptor modeling and ligand docking.

Methods

Ligand-Based Virtual Screening

A database SDF file was carefully built and characterized with basic unidimensional chemical descriptors, including 1622 compounds from our chemical library, with the Chemaxon suite of programs [13]. The database was curated with automated and manual filtering and the chemical diversity was assessed (data not shown).

A set of 3D conformers for each compound in the database was generated with OMEGA v2.2 [14] using the default settings, while for the query ligand, a single conformer was generated using more restrictive conditions [15]. It followed a shape-based overlay of conformers of each candidate molecule from the database to the query molecule with ROCS [14]. Default parameters were used, except for the minimum required Tanimoto shape overlap (*Tanimoto_cutoff*), which was set to 0.75. Shape overlays were refined with the so-called color force-field (*chemff ImplicitMillsDean*) [16], which scores the overlap of groups on the basis of their chemical properties (*-optchem flag*), using a Tanimoto-like scoring (known as *color score*). The method uses an implicit pK_a model, therefore avoiding the necessity of manual assessment of protonation states in our database. The Tanimoto and color scores are combined in a *combo score*, which in principle can adopt values from 0.7 (minimum accepted Tanimoto) to 2 (maximum value), and is used as scoring function in the ranking of compounds. The minimum accepted *combo score* was set to 1.2.

Binding Assays

The affinities of selected compounds for cloned human 5-HT_{2A} receptors were evaluated by *in vitro* binding assays

that used the radioligand [³H]-ketanserin according to previously described procedures [17]. K_i values expressed as pK_i were calculated according to the Cheng–Prusoff equation [18].

Homology Modeling and Ligand Docking

A homology model for the h5-HT_{2A} receptor was generated on the basis of the new crystal structure of β₂ adrenergic receptor [8]. This model was used for an automated docking exploration of compound **1** and clozapine, in order to determine their binding mode. The protocol for the generation of the receptor model and molecular docking is given in the supplementary material.

Results and Discussion

A chemical database was built including all the compounds present in the chemical library at Centro de Química (UM). The library was originally conceived to generate molecules with potential antioxidant and antimicrobial activity, and is actually composed of more than 1.500 compounds, where most of the heterocyclic scaffolds are combined with hydroxyl and phenolic substituents. The core structures of the molecules used in this study include mainly substituted purines and imidazo-pyridines (approximately 43% of the compounds tested). Substituted imidazoles, pyrimido-pyrimidines, pyrido-pyrimidines and a number of diaminomaleonitrile derivatives and nitrogen heterocycles including fused tricyclic structures were also used.

Our first concern was to identify the most suitable query molecule(s) for retrieving compounds with potential antipsychotic activity from our database by means of molecular similarity. From the pool of antipsychotics in clinical or preclinical stages, we first choose clozapine for two main reasons: i) it is still a gold standard in atypical antipsychotic activity, despite its side effects not related to its mechanism of action and ii) the molecule is chemically simple, with a minimum conformational flexibility due to its fused tricyclic structure. We have investigated the use of additional query compounds with similar pharmacological and chemical properties, but found no improvement in the hit identification results from independent or combined ROCS explorations (data not shown). Therefore, a single clozapine conformer constituted the query compound for a ROCS exploration of our database. The multiconformer database consisted of 46,959 conformers belonging to 1622 original compounds. 50 compounds were potentially rescued from the database, according to our filtering criteria. A good example of hit retrieving is shown in Figure 1, where a good overlay between the top ranked compound **1** and clozapine can be observed. Manual inspection followed in order to identify false positives, in particular compounds that were overestimated in the Tanimoto shape complementarity due to the consideration of highly un-

likely conformations of the compounds in the database. A possible solution to overcome this problem might be the use of more restrictive OMEGA parameters in the conformer generation, but this would clearly increase the risk of losing real hits, and manual post-processing of the ROCS results was preferred. It is worth to note that the default color forcefield [16] was accurate enough to retrieve hits, and there was no improvement in the results if we used the modifications proposed earlier for aminergic GPCRs ligands by Hawkins et al. [19]. This observation is in agreement with original remarks from that work, where the authors argue that the rationale behind the design of such modified forcefield was to emulate the restraints used by the docking methods (i.e., the interaction between the charged amino group and the conserved Asp3.32 in aminergic GPCRs) and had little effect on the results. Notably, the ROCS methodology showed an impressive performance precisely in the identification of known 5-HT_{2A} ligands from a database of decoys [19]. Moreover, since our database is mainly composed of neutral compounds lacking such amino group, the use of specific forcefield developed for charged compounds is precluded in the present case.

A total of seven compounds were selected for radioligand binding assays at h5-HT_{2A} receptor. These compounds are all members of the 9-methyl-purine family, with N6 substituent being either N-methyl piperazine or piperidine and an aryl derivative in position 2. Their preparation followed a new synthetic approach, where a substituted imidazole, obtained according to a previously described procedure [20] was combined with a substituted aldehyde under appropriate experimental conditions (unpublished results). The products were isolated in yields ranging from 40 to 65%. Table 1 shows these compounds with their experimental binding affinities and the scores obtained in the VS process. The pK_i of the query compound, clozapine, measured under the same experimental conditions was included for comparison [21]. The purine **1**, with a chlorine atom in the *meta* position of the phenyl group thus resembling clozapine (see fig. 1), presents a pK_i value of 5.53 ± 0.17. The most active compounds, **4** and **5**, with pK_i values of 5.81 ± 0.32 and 5.83 ± 0.13, respectively, bear electronegative groups in the *meta* and *para* positions of the phenyl group, indicating that this might be an important issue for the modulation of affinity. Finally, two compounds that lack a basic nitrogen (compounds **6** and **7**) show moderate affinities for the 5-HT_{2A} receptor. This result is in agreement with recent work reporting that a basic nitrogen is not mandatory for having affinity at these receptors, which opened the door for the design of new chemical compounds with lower affinity for antitargets, such as the IKr potassium channel [22].

A closer analysis of these results shows that we have been successful in the main goal of this study: to pick up original scaffolds suitable for the development of a new generation of compounds with antipsychotic activity. The

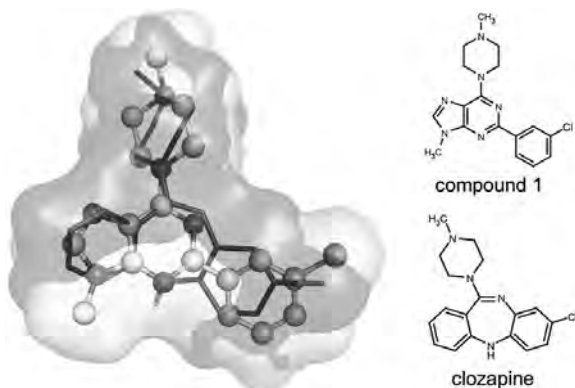


Figure 1. Predicted three-dimensional overlay of compound **1** (light gray, ball and sticks) and clozapine (dark gray, sticks). Connolly surface is displayed for each molecule as calculated with PyMOL (<http://www.pymol.org>)

Table 1. Hit compounds selected from the VS and confirmed by radioligand binding assays on the human 5-HT_{2A} receptor.

X	R ₁	Compound	pK _i (h5-HT _{2A})	Combo score
N-Me	Ph-3-Cl	1	5.53 ± 0.17	1.42
N-Me	Ph-2-Cl	2	5.08 ± 0.38	1.40
N-Me	Ph-4-Br	3	5.10 ± 0.54	1.40
N-Me	Ph-3,4-Cl ₂	4	5.81 ± 0.32	1.40
N-Me	Ph-4-Cl-3-CF ₃	5	5.83 ± 0.13	1.33
CH ₂ [a]	Ph-3-Cl	6	64.39 ± 3.03% [b]	1.27
CH ₂ [a]	Ph-2-Cl	7	64.91 ± 2.64% [b]	1.25
-	-	Clozapine	8.12 ± 0.07 [c]	query

[a] Compounds that lack a basic nitrogen;

[b] % of inhibition at 10 μM;

[c] Data from [21].

use of a 3D-LBVS technique such as ROCS methodology, based on shape matching and overlay of chemical groups with similar properties, was based on the ability to retrieve hits with unrelated chemistry to the query compounds [9] as well as the demonstrated success on hit identification for the particular case of GPCRs [19, 23]. As discussed above, the limited accuracy of homology models in VS and

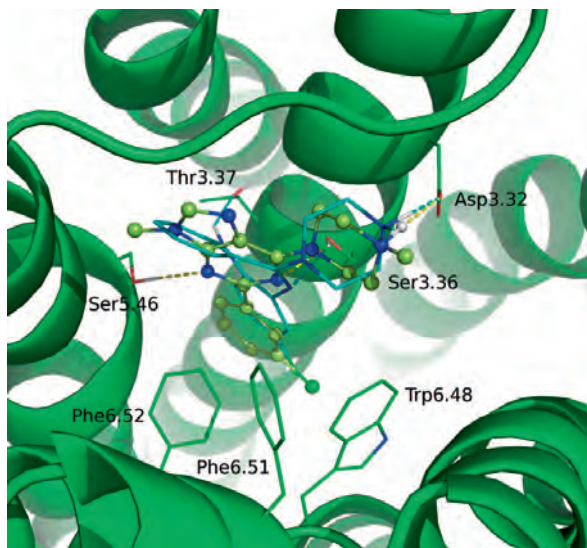


Figure 2. Binding mode proposed by molecular docking of clozapine (cyan, sticks) and compound **1** (yellow, ball and sticks) on the h5-HT_{2A} receptor model as suggested by docking exploration. Dashed lines indicate hydrogen bonds. Note that the superposition of the two molecules in the binding site is the same as the overlay presented in Fig. 1.

the polipharmacology of antipsychotics precluded the use of a RBVS approach. However, the initial series of hit compounds is still far from the nanomolar range in binding affinities at the 5-HT_{2A} receptor. In order to get a deeper understanding of the ligand-receptor interactions responsible to the modulation of binding affinity, we docked the query compound (clozapine) and the hit compound **1** to a h5-HT_{2A} receptor model (Fig. 2). As it can be appreciated in the figure, an independent docking search also identifies the same superposition between clozapine and the purine derivative as previously reported with the ROCS exploration. The binding mode of clozapine resembles the one proposed recently in an exhaustive docking exploration in several aminergic receptors [24], with the charged amino group interacting with the negatively charged Asp3.32, and the two aromatic rings located in the hydrophobic cavity of helix 6 (residues Phe6.51, Phe6.52 and Trp6.48) and helix 5, while the hydrogen bond of the N5 of clozapine is achieved with Thr3.37. As expected, the charged amino group of our hit compound is interacting through salt bridge with Asp3.32, and two additional hydrogen bonds are satisfied with Ser3.36 and Ser5.46 (this last residue interacts with clozapine in the model of Selent et al. [24]). However, it seems that the docking pose of **1** clashes with the helix 5 of the receptor, specially the methyl group at position 9 of the purine ring which is not well accommodated in the binding site. This might be a hypothesis for its relatively low affinity.

Conclusions

In this work we describe a new scaffold for aminergic receptors superimposable to clozapine as a first step for obtaining new possible clozapine-like purinergic derivatives active at serotonin 5-HT_{2A} receptors. The compounds have been retrieved from an original database not designed for aminergic GPCRs by a 3D-ligand based virtual screening tool and further confirmed by radioligand binding assays. A binding mode at the 5-HT_{2A} receptor that explains the bioactive molecular overlay with clozapine is presented, and the structural basis for its relatively low affinity is speculated. These findings are useful for the development of new scaffolds for antipsychotic drug discovery.

Acknowledgements

This work has been partially supported with funding from the Xunta de Galicia (Project PGIDIT07CSA003203PR). Mobility program from Spanish and Portuguese "Integrated Actions" has allowed the mobility of researchers between labs. H. G. T. is an Isidro Parga Pondal research Fellow and J. B. acknowledges Isabel Barreto Fellowship (both programs from Xunta de Galicia, Spain). D. R. is recipient of a predoctoral grant from the Fondo de Investigación Sanitaria (ISCIII, Spanish Ministry of Science and Innovation). C. C. and F. A. gratefully acknowledge Fundação para a Ciência e Tecnologia (Portugal) for Ph.D. (SFRH/BD/22270/2005) and Post-Ph.D. (SFRH/BPD/

26106/2005) grants, respectively. We thank the referee for invaluable comments that contributed to improve the manuscript.

References

- [1] WHO Fact Sheet Number 265, The World Health Organization, Geneva, Switzerland 2001.
- [2] J. A. Lieberman, C. A. Johns, J. M. Kane, K. Rai, A. V. Pisciotto, B. L. Saltz, A. Howard, *J. Clin. Psychiatry* **1988**, *49*, 271–277.
- [3] H. Y. Meltzer, S. Matsubara, J. C. Lee, *Pharmacol. Exp. Ther.* **1989**, *251*, 238–246.
- [4] B. L. Roth, D. J. Sheffler, W. K. Kroeze, *Nat. Rev. Drug Discovery*, **2004**, *3*, 353–359.
- [5] T. Lengauer, C. Lemmen, M. Rarey, M. Zimmermann, *Drug Discov. Today* **2004**, *9*, 27–34.
- [6] J. H. Nettles, J. L. Jenkins, A. Bender, Z. Deng, J. W. Davies, M. Glick, *J. Med. Chem.* **2006**, *49*, 6802–6810.
- [7] E. Gregori-Puigjane, J. Mestres, *J. Med. Chem.* **2006**, *46*, 1615–1622.
- [8] V. Cherezov, D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, R. C. Stevens, *Science* **2007**, *318*, 1258–1265.
- [9] T. Warne, M. J. Serrano-Vega, J. G. Baker, R. Moukhamet-zianov, P. C. Edwards, R. Henderson, A. G. Leslie, C. G. Tate, G. F. Schertler, *Nature* **2008**, *454*, 486–492.
- [10] J. H. Lange, J. H. Reinders, J. T. Tolboom, J. C. Glennon, H. K. Coolen, C. G. Kruse, *J. Med. Chem.* **2007**, *50*, 5103–5108.
- [11] S. Kortagere, W. J. Welsh, *J. Comput. Aided Mol. Des.* **2006**, *20*, 789–802.
- [12] R. Morphy, Z. Rankovic *Z. J. Med. Chem.* **2006**, *49*, 4961–70.
- [13] InstantJChem, Marvin and Calculator plugins, ChemAxon Ltd, Budapest, Hungary. <http://www.chemaxon.com>
- [14] OMEGA and ROCS, Openeye Scientific Software, Santa Fe, NM. <http://www.eyesopen.com>
- [15] J. Kirchmair, G. Wolber, C. Laggner, T. Langer, *J. Chem. Inf. Model.* **2006**, *46*, 1848–1861.
- [16] J. E. Mills, P. Dean, *J. Comput. Aided Mol. Des.* **1996**, *10*, 607–613.
- [17] J. Brea, J. Rodrigo, A. Carrieri, F. Sanz, M. I. Cadavid, M. J. Enguix, M. Villazón, G. Mengod, Y. Caro, C. F. Masaguer, E. Raviña, N. B. Centeno, A. Carotti, M. I. Loza, *J. Med. Chem.* **2002**, *45*, 54–71.
- [18] Y. Cheng, W. H. Prusoff, *Biochem. Pharmacol.* **1973**, *22*, 3099–3108.
- [19] P. C. D. Hawkins, A. G. Skillman, A. Nicholls, *J. Med. Chem.* **2007**, *50*, 74–82.
- [20] M. J. Alves, M. A. Carvalho, S. Carvalho, A. M. Dias, F. Fernandes, M. F. Proença, *Eur. J. Org. Chem.* **2007**, 4881–4887.
- [21] J. Brea, M. Castro, M. I. Loza, C. F. Masaguer, E. Raviña, C. Dezi, M. Pastor, F. Sanz, A. Cabrero-Castel, B. Galán-Rodríguez, E. Fernández-Espejo, R. Maldonado, P. Robledo, *Neuropharmacology* **2006**, *51*, 251–262.
- [22] T. Ladduwahetty, A. L. Boase, A. Mitchinson, C. Quin, S. Patel, K. Chapman, A. M. MacLeod, *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3201–3204.
- [23] S. W. Muchmore, A. J. Souers, I. Akritopoulou-Zanze, *Chem. Biol. Drug Des.* **2006**, *67*, 174–176.
- [24] J. Selent, L. López, F. Sanz, M. Pastor, *ChemMedChem* **2008**, *3*, 1194–1198.

Paper II

Michino M., Abola E., **GPCR Dock 2008 participants***, Brooks III C.L.,
Dixon J.S., Moulton J. and Stevens R.C.

Community-wide assessment of GPCR structure modelling and ligand
docking: GPCR Dock

Nat. Rev. Drug Discov, **2009** 8:455-63

* David Rodríguez is part of the consortium of participant researchers

Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008

Mayako Michino*, Enrique Abola*, GPCR Dock 2008 participants, Charles L. Brooks III†, J. Scott Dixon§, John Moul† and Raymond C. Stevens‡

Abstract | Recent breakthroughs in the determination of the crystal structures of G protein-coupled receptors (GPCRs) have provided new opportunities for structure-based drug design strategies targeting this protein family. With the aim of evaluating the current status of GPCR structure prediction and ligand docking, a community-wide, blind prediction assessment — GPCR Dock 2008 — was conducted in coordination with the publication of the crystal structure of the human adenosine A_{2A} receptor bound to the ligand ZM241385. Twenty-nine groups submitted 206 structural models before the release of the experimental structure, which were evaluated for the accuracy of the ligand binding mode and the overall receptor model compared with the crystal structure. This analysis highlights important aspects for success and future development, such as accurate modelling of structurally divergent regions and use of additional biochemical insight such as disulphide bridges in the extracellular loops.

* Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037, USA.

† Department of Chemistry and Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, USA.

‡ Daylight Chemical Information Systems Inc., Aliso Viejo, California 92656, USA.

§ Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, Rockville, Maryland 20850, USA.

¶ Departments of Molecular Biology and Chemistry, The Scripps Research Institute, La Jolla, California 92037, USA. Correspondence to R.C.S. e-mail: stevens@scripps.edu doi:10.1038/nrd2877

Published online 22 May 2009

Molecular modelling has an important role in rational drug design^{1,2}. Reliable three-dimensional models can provide valuable insights into basic principles of molecular recognition and aid in structure-based approaches to lead discovery and optimization³. G protein-coupled receptors (GPCRs) are membrane proteins involved in signal transduction pathways and are important therapeutic targets for numerous diseases^{4,5}. As such, significant structure prediction efforts using methods ranging from *de novo* to homology-based approaches have been applied to members of the GPCR family^{6,7}.

Until recently, most GPCR homology modelling efforts have been based on the templates of bovine rhodopsin and bacteriorhodopsin, with refinement of the models achieved through molecular dynamics simulations, ligand docking and incorporation of additional biochemical and biophysical data^{8–12}. The refinement step is necessary in building accurate models, especially around the ligand-binding site, owing to the expected structural differences among members of the family. These differences result from the generally low sequence identity and the large diversity of ligands accommodated within the family^{7,13–15}, and from the various conformational states that are associated with different levels of ligand efficacy^{16–18}.

The most recently solved GPCR structure is the 2.6 Å crystal structure of the human adenosine A_{2A} receptor bound to an antagonist¹⁹. Adenosine receptors belong to the class A rhodopsin-like GPCR family and represent promising therapeutic targets in a wide range of conditions, including cerebral and cardiac ischaemic diseases, sleep disorders, immune and inflammatory disorders, and cancer²⁰. The A_{2A} receptor structure shows an overall seven transmembrane (TM) helix architecture similar to that of the rhodopsin and adrenergic receptor structures, but with shifts in the positions and orientations of the helices and a markedly different structure of the extracellular loops¹⁹.

To evaluate current progress in GPCR structure prediction and the docking of potential ligands, as well as highlight areas for future efforts in method development, we carried out a community-wide, blind prediction assessment — GPCR Dock 2008 — in coordination with the publication of the human adenosine A_{2A} receptor structure in October 2008 (REF. 19). GPCR Dock 2008 was organized in a similar manner to the previous CASP (Critical Assessment of methods of Protein Structure) and CAPRI (Critical Assessment of Prediction of Interactions) studies^{21,22}. In this paper, we report the

Rhodopsin and bacteriorhodopsin

These two light-activated membrane proteins have a seven transmembrane alpha-helical bundle architecture that is similar to the general structure of the larger GPCR family.

outcome of the assessment together with our analysis of the current status of GPCR structure and ligand docking predictions.

GPCR Dock 2008

In August 2008, before the publication of the human adenosine A_{2A} receptor structure in October 2008 (REF. 19) and public release of the three-dimensional coordinates, participants were asked to predict and submit up to ten ranked models of the human A_{2A} receptor in complex

with the ligand ZM241385, starting from the amino acid sequence of the receptor and a two-dimensional structure of the ligand (see BOX 1 for list of GPCR Dock 2008 participants). A total of 63 different groups initially registered, with 206 models submitted by 29 different groups in the final data set (see [Supplementary information S1](#) (box) for details). Of the 206 submitted models, 37 were either missing the ligand or had incorrect bond connectivity for the ligand. We assessed the remaining 169 models for the prediction accuracy of the ligand binding mode, and all 206 models were assessed for the prediction accuracy of the receptor alone.

Box 1 | GPCR assessment participants

- Arthur Olson: Department of Molecular Biology, The Scripps Research Institute, USA
- Wiktor Jurkowski and Arne Elofsson: Center of Biomembrane Research, Department of Biochemistry & Biophysics, Stockholm University, Sweden
- Slawomir Filipek: Laboratory of Biomodelling, International Institute of Molecular and Cell Biology, Poland
- Irina Pogozheva and Andrei Lomize: Peptide Synthesis and Molecular Recognition Laboratory, University of Michigan, USA
- Bernard Maigret: Orpaillleur team, LORIA, Nancy University, France
- Jeremy Horst, Brady Bernard, Shyamala Iyer and Ram Samudrala: Computational Biology Group, University of Washington, USA; Ambrish Roy and Yang Zhang: Department of Molecular Biosciences, Center for Bioinformatics, University of Kansas, USA
- Osman Ugur Sezerman: Biological Science and Bioengineering, Sabanci University, Turkey
- Gregory V. Nikiforovich: MolLife Design LLC, USA; Christina M. Taylor: Department of Biochemistry and Molecular Biophysics, Washington University, USA
- Stefano Costanzi: Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, USA
- Y. Vorobjev, N. Bakulina, and V. Solovjev: Department of Computer Science, Royal Holloway, University of London and Softberry Inc., UK
- Kazuhiko Kanou, Daisuke Takaya, Genki Terashi, Mayuko Takeda-Shitaka and Hideaki Umeyama: School of Pharmacy, Kitasato University and RIKEN Systems and Structural Biology Centre, Japan
- William A. Goddard III, Youyong Li, Soo-Kyung Kim, Bartosz Trzaskowski, Ravinder Abrol and Adam Griffith: Materials and Process Simulation Center, California Institute of Technology, USA
- Vsevolod Katritch, Manuel Rueda and Ruben Abagyan: Molsoft LLC, USA
- Ian Davis, Patrick Barth and David Baker: Department of Biochemistry, University of Washington, USA
- Michael Feig: Department of Biochemistry and Molecular Biology, Michigan State University, USA
- Michal Brylinski, Hongyi Zhou, Seung Yup Lee and Jeffrey Skolnick: Center for the Study of Systems Biology, Georgia Institute of Technology, USA
- Liliana Ostropovici-Halip and Cristian Bologa: Division of Biocomputing, University of New Mexico, USA
- Polo Lam and Ruben Abagyan: Department of Molecular Biology, The Scripps Research Institute, USA
- Eric S. Dawson, Kristian Kaufmann, Nils Woetzel and Jens Meiler: Center for Structural Biology, Vanderbilt University, USA
- Feng Ding, Adrian Serohijos, Shuangye Yin and Nikolay V. Dokholyan: Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, USA
- David Rodriguez and Hugo Gutiérrez-de-Terán: Fundación Pública Galega de Medicina Xenómica, Complejo Hospitalario Universitario de Santiago de Compostela, Spain
- Henri Xhaard: Center for Drug Research, Faculty of Pharmacy, University of Helsinki, Finland

For full details, see [Supplementary information S1](#) (box).

Assessment criteria. Assessment criteria are dependent on the purpose of the generated models. Given the value of the GPCR structural models in expanding our knowledge in basic molecular recognition and their potential use in the design and development of new small molecules, the quality of the models was primarily assessed by the accuracy of the ligand binding mode. Particular attention was given to the fact that the crystal structure is a static structure with positional errors, and the value of modelling is ultimately to guide drug discovery and provide biological insight. Our numerical measure of accuracy for the ligand binding mode was based on two metrics: ligand root mean square deviation (RMSD) and the number of correct receptor–ligand contacts. Neither metric alone was sufficient to capture the accuracy of prediction around the ligand binding site; hence, both were used and combined into a z-score to rank the models.

The ligand RMSD between the model and the crystal structure was calculated as the coordinate RMSD for the 25 non-hydrogen atoms of ZM241385 after superimposing the C α atoms of the protein in the model and the crystal structure. In addition, the ligand RMSD is also calculated excluding the phenoxy group of ZM241385 that has high B-factor values. The number of correct contacts is counted as the number of correctly predicted native contacts observed between protein atoms and the ligand. A native contact is defined as any interatomic distance within 4 Å of the ligand in the crystal structure. There are 75 such receptor–ligand contacts, and an additional 15 contacts formed with water.

The models were ranked by assigning a combined mixed z-score to each model. The combined z-score was calculated as the average of z-scores for ligand RMSD and the number of correct contacts:

$$Z_{\text{combined}} = (-Z_{\text{ligand RMSD}} + Z_{\text{Number of correct contacts}})/2.$$

The z-scores for ligand RMSD and the number of correct contacts were computed by the following steps. First, a z-score was assigned to each model using the average and standard deviation (SD) values from all models. Second, the average and SD was re-computed excluding models with z-scores that were more than two SDs above (for ligand RMSD) or below (for the number of correct contacts) the average. Third, a z-score was reassigned to each model using the revised average and SD values obtained in step two. The best model — that is, the model with the highest combined z-score — from each group was analysed.

Molecular dynamics simulation

This molecular modelling approach uses numerical integration to solve the equations of motion based on the forces arising from interatomic interactions. The dynamic behaviour of atoms in a macromolecular system, such as that in a membrane protein, can be understood by running a molecular dynamics (MD) simulation. MD simulation can also be used to refine structural models of proteins and protein–ligand complexes.

Ligand docking

A molecular modelling approach that predicts the ligand binding mode within a targeted binding site. In this approach, the known or predicted three-dimensional structure of a protein is probed using computationally generated energy landscapes to identify the most favourable binding pose for the ligand.

RMSD (root mean square deviation)

RMSD is used as a quantitative measure of the similarity between two superimposed atomic coordinates. RMSD values (units of Å) can be calculated for any type and subset of atoms; for example, Ca atoms of proteins (Ca RMSD) for all residues, for residues in the transmembrane helices or the loops; heavy atoms of small-molecule ligands (ligand RMSD).

Z-score

A standard dimensionless score that normalizes a value with respect to the sample mean and standard deviation.

Ca atoms

The chiral carbon atoms to which the primary amine, the carboxylic group and the side chain are attached to in an amino acid. Comparison of three-dimensional structures of proteins is sometimes carried out by superimposing the Ca atoms of proteins as this provides a simple estimate of the similarity of their skeleton or backbone structure.

Overall outcome of analysis. The submitted models showed a wide distribution in prediction accuracy of the ligand binding mode, with average values of 9.5 Å (SD 3.8 Å) for ligand RMSD (FIG. 1a) and 4 (SD 7) for the number of correct contacts. These statistics indicate that the majority of the submitted models did not predict the ligand position and the binding interactions very accurately. The lack of a strong correlation between ligand RMSD and binding site RMSD (FIG. 1b) (for example, models with a binding site RMSD of less than 4.0 Å have a range of 2.8 to 17.2 Å ligand RMSD), suggests that the performance of some ligand docking methods can be improved.

Very few models score well in both ligand RMSD and the number of correct contacts (only 13 out of the 169 total receptor–ligand models have a combined z-score greater than 1, compared with 40 models that score well solely in ligand RMSD ($Z_{\text{ligand RMSD}}$ less than -1 Å). For models with relatively low ligand RMSD values but a small number of correct contacts, the inaccuracy in binding interactions could be attributed to errors in the side chain placement of the ligand binding residues. Although nearly a third of the models capture the hydrogen bonding interaction between the N253^{6,55} side chain and the exocyclic N15 atom of the ligand (44 out of 169 models have a N253 OD1–ZM241385 N15 interaction distance of less than 4 Å), other key receptor–ligand interactions, such as the aromatic stacking interaction between the F168^{5,29} side chain and the bicyclic ring of the ligand, are not captured well in most models (FIG. 2).

Although the overall outcome clearly shows that there are remaining challenges in accurately predicting the ligand binding mode, the quality of the predictions for the receptor alone seem relatively good: 4.2 ± 0.9 Å for the receptor Ca RMSD, and 2.8 ± 0.5 Å for the TM helices Ca RMSD. Not surprisingly, loop regions, with the exception of the short intracellular loop 1 (ICL1), are not modelled very well in most of the models (FIG. 3a,b and FIG. 4a,b). It is notable that some groups that accurately predicted the TM region of the receptor did not predict the ligand binding mode very well (for example, the TM Ca RMSD is 2.0 Å for the model submitted by I. Pogozheva and A. Lomize, and 2.1 Å for the model submitted by J. Horst and A. Roy), indicating that the methods for modelling the receptor and docking of the ligand can be generally considered as distinct steps in the generation of models for the receptor–ligand complex.

Analysis of the best models

Despite the challenges in accurately predicting the receptor–ligand interactions, some models had consistent features with the crystal structure, although model ranking continues to be one of the most challenging areas of development. Here, we focus on the predictions from the top ten groups, ranked according to the combined z-score, and assess the model quality in greater detail (FIG. 4c). Note that, with predictions for only one target, the statistical significance of the group ranking cannot be judged as is typically done in CASP experiments by a head-to-head comparison of common targets between

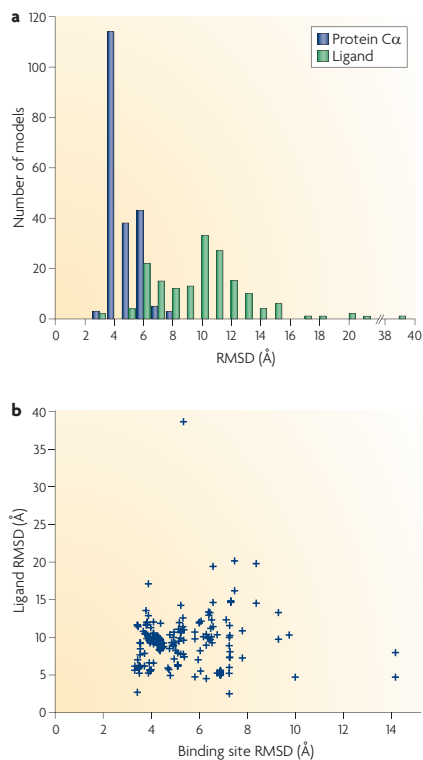


Figure 1 | Root mean square deviation (RMSD) of submitted models. **a** | Distribution of ligand RMSD (green bars) and protein Ca RMSD (blue bars) for all models. **b** | A scatterplot of ligand RMSD (y axis) versus binding site RMSD (x axis) for all models. The binding site RMSD values are calculated for heavy atoms of the binding site residues (F168^{5,29}, E169^{5,30}, M177^{5,38}, W246^{6,48}, L249^{6,51}, H250^{6,52}, N253^{6,55}, H264^{6,66}, M270^{7,35}) after the models were superimposed to the crystal structure using the Ca atoms of the protein.

the top groups²³. To support our selection of the best predictions, we ranked all models using an alternative metric — binding site contact RMSD — which gives all ligand binding residues equal weight and is an RMSD of receptor–ligand contact distance for all ligand-binding residues. We found that both the z-score ranking and the contact RMSD ranking agree on the selection of the best model.

The best model overall (submitted by S. Costanzi) has a ligand RMSD of 2.8 Å and 34 of 75 correct contacts (FIG. 5a and TABLE 1). The ligand is modelled in a native-like binding pose, with an extended conformation and a nearly perpendicular orientation to the membrane plane. The model accurately predicts some of the key receptor–ligand interactions: it captures the hydrogen

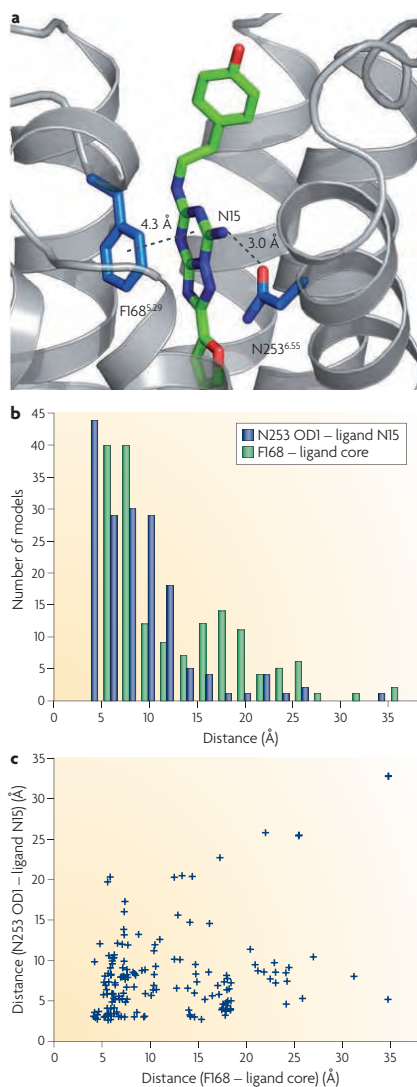


Figure 2 | Statistics of the two key receptor-ligand interactions in all models. **a** | The hydrogen bonding interaction with N253^{6,55} and the aromatic stacking interaction with F168^{5,29} are shown by dashed lines with the distance measurements from the crystal structure. **b** | Distribution of the distance between the side chain carbonyl oxygen OD1 atom in N253^{6,55} and the exocyclic N15 atom of the ligand (ZM241385), and the average interatomic distance for the aromatic stacking interaction between the heavy atoms in the F168^{5,29} side chain and the bicyclic ring (atoms C11, N12, N13, C14, N15, N16, N17, C18, N19, C20) of the ligand. **c** | A scatterplot of the distances for the hydrogen bonding interaction (y axis) versus the aromatic stacking interaction (x axis).

B-factor

A descriptor that reflects the fluctuation of atomic position from an atom's average position and provides important insight into a protein's potential dynamic behaviour.

Hydrogen bond

Attractive interaction between one electronegative atom and a hydrogen covalently bonded to another electronegative atom such as nitrogen or oxygen.

Aromatic stacking

Attractive interactions between the aromatic rings of amino acids. Overlapping of p-orbitals of π -conjugated systems result in the rings arranging themselves in preferred orientations.

bonding interaction between the N253^{6,55} side chain and the exocyclic amino group (N15 atom) of the ligand, and the aromatic stacking interaction between the F168^{5,29} side chain and the bicyclic triazolotriazine core of the ligand. Compared with the crystal structure, the ligand in the model is positioned deeper in the binding pocket, bringing the furan ring closer to TM helices III and V. The inaccuracy in the ligand position is most probably due to errors in the side chain positions of the two crucial ligand binding residues (F168^{5,29} and E169^{5,30}) in extracellular loop 2 and the side chain orientation of M177^{5,38} at the extracellular end of TM helix V. The aromatic ring of F168^{5,29}, which interacts with the bicyclic ring, is positioned too deeply; the adjacent E169^{5,30} forms a hydrogen bonding interaction with the hydroxyl group in the phenolic substituent, instead of the exocyclic N15 atom near the bicyclic ring; and the side chain of M177^{5,38} is not oriented towards the binding cavity. In addition, the family conserved disulphide bond between C77^{3,25}-C166^{5,27} is predicted accurately, but the disulphide bond in extracellular loop 3 between C259^{6,61}-C262^{6,64} is not, presumably contributing to the inaccuracy in the side chain orientation of H264^{6,66}, which is not pointed towards the binding site.

The best predictions from the top six groups (S. Costanzi; V. Katritch and R. Abagyan; P. Lam and R. Abagyan; I. Davis, P. Barth and D. Baker; B. Maigret; W. Jurkowski and A. Elofsson) highlight the successes and challenges in accurately predicting the ligand binding pose and receptor-ligand interactions (FIG. 5B, C, D and TABLE 1). The extended ligand conformation is accurately predicted in all six models, and the nearly perpendicular orientation is captured in four of the six models. The hydrogen bonding interaction between the N253^{6,55} side chain and the exocyclic N15 atom of the ligand is correctly modelled in four models; however, in one of the four, the ligand makes no interaction with residues in extracellular loop 2. The aromatic stacking interaction between the F168^{5,29} side chain and the bicyclic ring of the ligand is correctly modelled in four models; however, in all four models, the ligand is positioned too deeply in the binding pocket, and the M177^{5,38} side chain is not oriented towards the binding cavity. There is one model that does not accurately capture either the hydrogen bonding interaction with N253^{6,55} or the aromatic stacking interaction with F168^{5,29}, whereas five of the six models accurately predict the family conserved disulphide bond between C77^{3,25}-C166^{5,27}. None of the six models capture the hydrogen bonding interaction between E169^{5,30} in extracellular loop 2 and the exocyclic N15 atom of the ligand.

Other models that ranked near the top (those submitted by K. Kanou, W. A. Goddard, C. Bologna and A. Olson) are slightly less accurate, but show similar trends to the top six models in their ability to accurately predict the ligand binding mode (TABLE 1). The ligand is modelled in a native-like extended conformation in three of the four models. The hydrogen bonding interaction between the N253^{6,55} side chain and the exocyclic N15 atom of the ligand is modelled accurately in three of the four models, whereas the aromatic stacking interaction

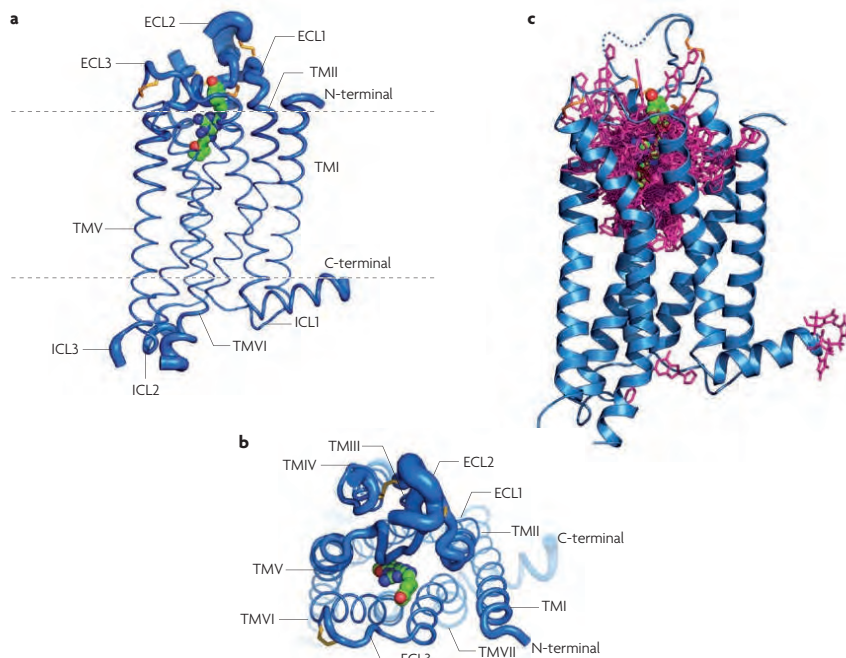


Figure 3 | Superposition of all 206 submitted models to the crystal structure of the human adenosine A_{2A} receptor. Protein Ca atom superposition between each model and the crystal structure was done using the align command in PyMOL (version 1.0r2, www.pymol.org) (Protein Data Bank ID code: 3EML without the T4-lysozyme). **a,b** | The receptor is shown as two orthogonal views of Ca traces, with tube thickness being proportional to the root mean square deviation (RMSD) of each Ca position, showing how well the transmembrane (TM) regions were modelled and how much uncertainty there is in the loop regions. **c** | A superposition of stick diagrams of the ligand (ZM241385) from 169 models; a CPK model is used to delineate the observed position in the crystal structure. The carboxy-terminus (residue numbers greater than 306) is removed from all models. ECL, extracellular loop; ICL, intracellular loop.

between the F168^{5,29} side chain and the bicyclic ring of the ligand is modelled accurately in only one of the four models. The family conserved disulphide bond between C77^{3,25}–C166^{5,27} is captured in two models. Remarkably, one of the models (submitted by W. A. Goddard) accurately places the E169^{5,30} side chain proximal to the exocyclic N15 atom of the ligand, and almost captures the hydrogen bonding interaction, even though the overall conformation of extracellular loop 2 is inaccurate.

The best predictions were generally not ranked as the best models by the predictors at the time of model submission (before the release of the crystal structure) (TABLE 1). Only two of the six best models were ranked first, and three of the six groups show a weak correlation between their model ranking and the model quality as assessed by the combined z-score for the accuracy around the ligand-binding site. Furthermore, the additional models submitted by the six groups are generally of lower quality than the best predictions (TABLE 1). Only one of the six best models has a z-score that is within one SD of the group average z-score.

Status of GPCR structure modelling and docking

The assessment of the submitted models showed that the best participating methods have the ability to predict close, native-like ligand binding, but have limitations in capturing all of the key receptor–ligand interactions and correctly estimating model quality by ranking. The majority of the submitted models are quite far from predicting a native-like ligand binding pose. The most challenging aspect of GPCR structure prediction highlighted in this assessment seems to be in accurately modelling the ligand interactions with residues in the extracellular loop regions. This result is not surprising given the lack of structural homology in the loops among the known GPCR structures²⁴, and the general difficulties in modelling loop regions^{25,26}.

The most successful prediction methods relied on homology modelling approaches based on the template structures of β -adrenergic receptors, and in some cases with the additional template structures of rhodopsin (Protein Data Bank ID code: 2RH1 (β_2 AR), 2VT4 (β_1 AR), 1U19 (bovine rhodopsin), 2Z73 (squid rhodopsin)) to

generate models of the receptor, followed by docking of the ligand to one or more receptor models using small-molecule docking programmes such as Glide²⁷, ICM²⁸, GOLD²⁹ and AutoDock³⁰ (see [Supplementary information S1](#) (box) for description of prediction methods). The alignment of the human A_{2A} receptor sequence to the

template structure seemed to have been straightforward, given the family conserved motifs and residues in the TM helices³¹. The extracellular loop 2 was modelled by *de novo* approaches in many of the top predictions (V. Katritch and R. Abagyan; P. Lam and R. Abagyan; I. Davis, P. Barth and D. Baker; W. Jurkowski and A. Elofsson; W. A. Goddard), but only partially modelled in the best prediction (S. Costanzi) for a short segment of eight residues, located amino-terminal to TM helix V, which includes the disulphide bond forming C166^{5,27}. Some of the criteria used to select and rank the final receptor–ligand complex models were: docking scores, conformational energy of the complex, agreement with mutagenesis and structure–activity relationship data, and binding selectivity studied by virtual ligand screening or by modelling other subtypes of adenosine receptor.

The reliability of the homology modelling approach depends on the availability of suitable templates³². The results of the current assessment show that the structures of β -adrenergic receptors alone or together with rhodopsin were suitable transmembrane templates in predicting the general structure of the adenosine A_{2A} receptor. However, given the expected structural diversity in class A GPCRs, it is unclear whether the current set of techniques applied to the structure prediction of the A_{2A}-ZM241385 complex would result in a similar level of accuracy for the prediction of other GPCRs, especially for those belonging to subfamilies that are phylogenetically distant from the amine and the opsin receptor clusters³³. We believe the database of GPCR structures needs to expand further to provide suitable templates for accurate modelling of those other receptors.

The inaccuracies in homology models can arise from errors in side chain packing, main chain shifts in aligned regions, errors in unaligned loop regions, misalignments and incorrect templates³⁴. These errors relate to the issue of ‘adding value’ to the template structure, which was addressed in the recent CASP experiment³⁵, and also seems to be applicable to GPCR modelling. Indeed, ligand interactions with residues located in structurally divergent regions from the templates are consistently not modelled accurately in all of the six best predictions: the hydrogen bonding interaction between E169^{5,30} in extracellular loop 2 and the exocyclic N15 atom of the ligand is not captured, and the side chains of H264^{6,66} in extracellular loop 3 and M177^{5,38} in the extended bulge structure unique to the A_{2A} receptor at the extracellular end of TM helix V are not oriented towards the binding site. An exception is the aromatic stacking interaction between F168^{5,29} in extracellular loop 2 and the bicyclic ring of the ligand, which is correctly modelled in some of the predictions. F168^{5,29} is located in the loop, but it is structurally homologous to F193^{5,32}, which interacts with the carbazole heterocycle of the ligand carazolol in the β_2 AR structure; hence modelling of this interaction may have been guided by homology. Interestingly, F168^{5,29} is modelled more accurately than E169^{5,30} even though mutagenesis data showed that mutation of E169^{5,30} to alanine reduces the affinity for both antagonists and agonists³⁶, and no data is available for F168^{5,29}.

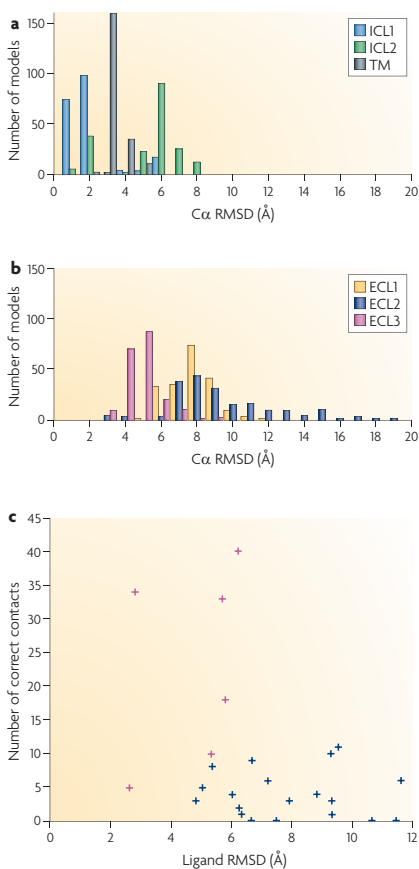


Figure 4 | Model analysis. **a, b** | Distribution of C α root mean square deviations (RMSDs) for adenosine A_{2A} receptor domains. Panel **a** shows C α RMSDs for transmembrane (TM) helices and intracellular loop (ICL) regions: helix I: 6–34; helix II: 40–67; helix III: 73–107; helix IV: 117–142; helix V: 173–205; helix VI: 222–258; helix VII: 266–291; ICL1: 35–39; ICL2: 108–116. Panel **b** shows C α RMSDs for extracellular loop (ECL) regions: ECL1: 68–72; ECL2: 143–172 (excluding 149–155 that are missing in the crystal structure); ECL3: 259–265. **c** | Scatterplot of the number of correct contacts versus ligand RMSDs for the best predictions from all groups. The best predictions from the top six groups are marked as mauve crosses.

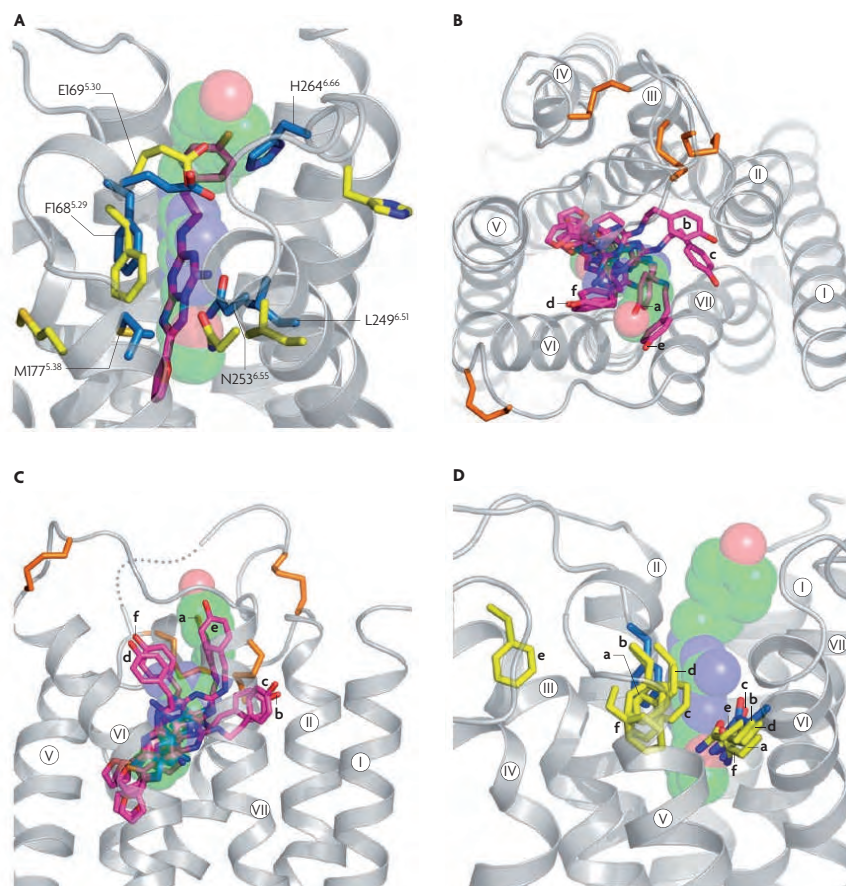


Figure 5 | Comparison between the best models and the crystal structure around the ligand-binding site.

The ligand and the ligand-binding residues F168^{5,29}, E169^{5,30}, M177^{5,38}, L249^{6,51}, N253^{6,55} and H264^{6,66} are shown for the best model (S. Costanzi) and the crystal structure (A). The ligand is shown as magenta sticks for the model, and as green semitransparent spheres for the crystal structure; the ligand-binding residues are shown as yellow sticks for the model, and blue sticks for the crystal structure. Extracellular (B) and side views (C) of the ligand in the binding pocket for the best predictions from the top six groups (magenta sticks for models and green spheres for the crystal structure). The receptor crystal structure is shown as grey ribbons. The disulphide bonds are shown as orange sticks. D | The ligand-binding residues F168^{5,29} and N253^{6,55} are shown as sticks for the best predictions from the top six groups (yellow for models and blue for the crystal structure). In B–D the models are labelled as: a, S. Costanzi; b, V. Katritch; c, P. Lam; d, I. Davis; e, B. Maigret; f, W. Jurkowski.

The inaccuracy in the orientation of the ligand binding pose — for example, the parallel orientation with the phenolic substituent positioned close to TM helices II and III — may in part be due to the inaccurate modelling of the helical shifts in TM helices I, II and III. The helical shifts alter the location of the binding pocket and redefine the pocket size and shape¹⁵; thus, it is expected that accurately modelling the helical shifts would contribute to a better prediction of the

ligand binding pose. The helical shifts were most accurately modelled by an effective use of multiple template structures of rhodopsin and β -adrenergic receptors (I. Pogozheva and A. Lomize), or an all-atom refinement approach implemented by the ROSETTA programme using a physically realistic model that recapitulated protein interatomic and protein–solvent interactions in the membrane environment³⁷ (I. Davis, P. Barth and D. Baker).

Table 1 | Summary of results for the best models from the top ranking groups

Group name	Rank (total number of models)	Ligand RMSD (Å)	Ligand RMSD without phenoxy group (Å)	Number of correct contacts	Binding site residues RMSD (Å)	Protein C α RMSD (Å)	TM I–VII C α RMSD (Å)	ECL2 C α RMSD (Å)	Combined z-score (average \pm SD)
Costanzi	2 (4)	2.8	2.7	34	3.4	3.0 (266)	2.5 (212)	3.8 (8)	3.02 (0.86 \pm 1.48)
Katritch & Abagyan	1 (10)	6.2	4.0	40	3.5	4.0 (283)	2.7 (214)	8.9 (23)	2.76 (1.89 \pm 1.13)
Lam & Abagyan	1 (3)	5.7	3.6	33	3.3	4.1 (283)	3.6 (214)	7.3 (23)	2.42 (0.88 \pm 1.34)
Davis, Barth & Baker	4 (5)	5.8	5.4	18	4.0	3.5 (283)	2.1 (214)	8.4 (23)	1.46 (0.16 \pm 0.86)
Maignret	8 (10)	2.6	2.1	5	7.3	5.1 (283)	4.1 (214)	9.1 (23)	1.23 (0.05 \pm 0.57)
Jurkowski & Elofsson	2 (8)	5.3	5.2	10	3.9	6.2 (283)	2.9 (214)	12.7 (23)	1.04 (−0.02 \pm 0.98)
Kanou	7 (10)	5.4	5.5	8	6.9	3.5 (279)	2.8 (214)	7.1 (23)	0.91 (0.66 \pm 0.11)
Goddard	8 (10)	5.0	3.9	5	4.8	4.3 (284)	2.5 (214)	10.7 (23)	0.78 (0.16 \pm 0.37)
Bologa	3 (10)	6.7	2.8	9	3.9	3.4 (278)	2.5 (213)	7.2 (19)	0.72 (−0.14 \pm 0.39)
Olson	1 (9)	4.8	4.7	3	5.8	3.5 (284)	2.3 (214)	7.5 (23)	0.69 (−0.14 \pm 0.58)

Participants were allowed to submit up to 10 models. Rank indicates the ranking that the participant assigned to their best model as determined in the GPCR Dock 2008 study with the total number of models submitted by that participant in parentheses. The root mean square deviation (RMSD) values were calculated for the heavy atoms of the ligand ZM241385 (all 25 atoms and partially without the phenoxy group), heavy atoms of the binding site residues (F168⁵²⁹, E169³⁰, M177⁵³, W246⁴⁸, L249⁵¹, H250⁵², N253⁵⁵, H264⁶⁶, M270³⁵), C α atoms of all residues, C α atoms of residues in the transmembrane helices (TM) I to VII (helix I: 6–34; helix II: 40–67; helix III: 73–107; helix IV: 117–142; helix V: 173–205; helix VI: 222–258; helix VII: 266–291), and C α atoms of residues in extracellular loop 2 (ECL2) (143–172 excluding 149–155 that are missing in the crystal structure). All RMSD values were obtained after the models were superimposed to the crystal structure, using the protein C α atoms in PyMOL (version 1.0r2, www.pymol.org). The assignment of residues in the ligand-binding site and the secondary structure elements is from the Protein Data Bank header section (PDB ID: 3EML). The number of residues used in the RMSD calculation is in brackets. The combined z-score value for the best model, as well as the average and standard deviation (SD) values for all models submitted by each group, are shown.

Other sources of error include not modelling the water molecules that are either structurally important or directly involved in ligand binding interactions⁵. The ligand binding cavity in the A_{2A}-ZM241385 structure has four ordered water molecules¹⁹, yet none of the submitted predictions included water molecules. We tried re-docking the ligand to the crystal structure using ICM²⁸ and found that a native-like binding pose (within 1 Å heavy atom RMSD for the bicyclic ring and the furanyl substituent of the ligand, and less than 3 Å overall ligand RMSD) can be recovered without any water molecules, which suggests that water may not be critical for accurately predicting the ligand interactions. However, modelling water molecules together with the ligand might contribute to a better prediction of the ligand binding pose or affinity. Additional re-docking studies with the docking protocols used by the participating methods would help assess the effect of the water molecules, and the accuracy of the docking methods separately from that of the receptor modelling methods.

Finally, it is interesting that the best model was from the S. Costanzi group, which has previously worked on adenosine receptor modelling and docking. Their

domain knowledge on the adenosine receptor is likely to have been crucial for the evaluation and interpretation of the mutagenesis and ligand interaction data.

Conclusions

Accurate prediction of GPCR structure and ligand interactions remains a challenge, and the approach will improve with the recent availability of experimentally solved GPCRs. Assessment of these predictions highlights similar issues addressed by the CASP predictions for template-based modelling targets; that is, the difficulty in loop modelling, refinement and improvement over the best available template and model ranking. Accurate modelling of the structurally divergent regions (such as the extracellular loops that form defined architectures), and disulphide bond formation affecting helix residue registry and helical shifts in the TM region seem to be crucial for accurately predicting the key ligand interactions in GPCRs, and this area is perhaps the most in need of technological development. Progress in GPCR modelling and docking will require further improvements in the current prediction methods to enhance the best available templates and generate models that will be more useful for applications in structure-based drug design.

- Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **303**, 1813–1818 (2004).
- Richon, A. B. Current status and future direction of the molecular modeling industry. *Drug Discov. Today* **13**, 665–669 (2008).
- Kitchen, D. B., Decornez, H., Furr, J. R. & Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Rev. Drug Discov.* **3**, 935–949 (2004).
- Drews, J. Drug discovery: a historical perspective. *Science* **287**, 1960–1964 (2000).
- Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **3**, 928–944 (2002).
- Becker, O. M. *et al.* G protein-coupled receptors: *in silico* drug discovery in 3D. *Proc. Natl Acad. Sci. USA* **101**, 11304–11309 (2004).
- Ballesteros, J. & Palczewski, K. G protein-coupled receptor drug discovery: implications from the crystal structure of rhodopsin. *Curr. Opin. Drug Discov. Devel.* **4**, 561–574 (2001).
- Bu, L., Michino, M., Wolf, R. M. & Brooks, C. L. III. Improved model building and assessment of the Calcium-sensing receptor transmembrane domain. *Proteins* **71**, 215–226 (2008).
- Henin, J. *et al.* Probing a model of a GPCR/ligand complex in an explicit membrane environment: the human cholecystokinin-1 receptor. *Biophys. J.* **90**, 1232–1240 (2006).
- Fowler, C. B., Pogozeva, I. D., LeVine, H., 3rd & Mosberg, H. I. Refinement of a homology model of the mu-opioid receptor using distance constraints from intrinsic and engineered zinc-binding sites. *Biochemistry* **43**, 8700–8710 (2004).
- Evers, A. & Klabunde, T. Structure-based drug discovery using GPCR homology modeling: successful virtual screening for antagonists of the alpha 1A adrenergic receptor. *J. Med. Chem.* **48**, 1088–1097 (2005).
- Manivet, P. *et al.* The serotonin binding site of human and murine 5-HT_{2B} receptors: molecular modeling and site-directed mutagenesis. *J. Biol. Chem.* **277**, 17170–17178 (2002).
- Archer, E., Maigret, B., Escrieut, C., Pradayrol, L. & Fourmy, D. Rhodopsin crystal: new template yielding realistic models of G-protein-coupled receptors? *Trends Pharmacol. Sci.* **24**, 36–40 (2003).
- Gershengorn, M. C. & Osman, R. Minireview: Insights into G protein-coupled receptor function using molecular models. *Endocrinology* **142**, 2–10 (2001).
- Ballesteros, J. A., Shi, L. & Javitch, J. A. Structural mimicry in G protein-coupled receptors: implications of the high-resolution structure of rhodopsin for structure-function analysis of rhodopsin-like receptors. *Mol. Pharmacol.* **60**, 1–19 (2001).
- Kobilka, B. K. & Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28**, 397–406 (2007).
- Bhattacharya, S., Hall, S. E., Li, H. & Vaidehi, N. Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation. *Biophys. J.* **94**, 2027–2042 (2008).
- Kenakin, T. Efficacy at G-protein-coupled receptors. *Nature Rev. Drug Discov.* **1**, 103–110 (2002).
- Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322**, 1211–1217 (2008).
- The human adenosine A_{2A} receptor crystal structure served as the experimental template for comparison for this modelling and docking assessment. This is the second human GPCR structure to be experimentally determined.**
- Jacobson, K. A. & Gao, Z. G. Adenosine receptors as therapeutic targets. *Nature Rev. Drug Discov.* **5**, 247–264 (2006).
- Moult, J. *et al.* Critical assessment of methods of protein structure prediction-Round VII. *Proteins* **69** (Suppl. 8), 3–9 (2007).
- The very successful CASP (Critical Assessment of Protein Structure) project started in 1994 and served as the model to conduct the reported GPCR Dock 2008 modelling and docking assessment.**
- Lensink, M. F., Mendez, R. & Wodak, S. J. Docking and scoring protein complexes: CAPRI 3rd Edition. *Proteins* **69**, 704–718 (2007).
- Kopp, J., Bordoli, L., Battay, J. N., Kiefer, F. & Schwede, T. Assessment of CASP7 predictions for template-based modeling targets. *Proteins* **69** (Suppl. 8), 38–56 (2007).
- Kobilka, B. & Schertler, G. F. New G-protein-coupled receptor crystal structures: insights and limitations. *Trends Pharmacol. Sci.* **29**, 79–83 (2008).
- Jacobson, M. P. *et al.* A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367 (2004).
- Rohl, C. A., Strauss, C. E., Chivian, D. & Baker, D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* **55**, 656–677 (2004).
- Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749 (2004).
- Totrov, M. & Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins* **29** (Suppl. 1), 215–220 (1997).
- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **52**, 609–623 (2003).
- Morris, G. *et al.* Automated docking using a lamarkian genetic algorithm and empirical binding free energy function. *J. Comput. Chem.* **19**, 1639–1662 (1998).
- Mirzadegan, T., Benko, G., Filipek, S. & Palczewski, K. Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* **42**, 2759–2767 (2003).
- Baker, D. & Sali, A. Protein structure prediction and structural genomics. *Science* **294**, 93–96 (2001).
- The use of protein models and docking is dependent on how such data will be used. In this paper, Baker and Sali provide an excellent presentation of where models are useful, in particular as hypothesis generators with the application being dependent on the resolution of the structure.**
- Fredriksson, R., Lagerstrom, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogue groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Marti-Renom, M. A. *et al.* Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
- Read, R. J. & Chavali, G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. *Proteins* **69** (Suppl. 8), 27–37 (2007).
- Kim, J. *et al.* Glutamate residues in the second extracellular loop of the human A2a adenosine receptor are required for ligand recognition. *Mol. Pharmacol.* **49**, 683–691 (1996).
- Barth, P., Schonbrun, J. & Baker, D. Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc. Natl Acad. Sci. USA* **104**, 15682–15687 (2007).

Acknowledgements

We thank M. Hanson, V.-P. Jaakola, C. Roth and V. Cherezov for help with the analysis and comments on the manuscript, and K. Kadyshkevskaya and V. Cherezov for figure preparation. We are grateful to the Goddard group for providing the script to calculate the binding site contact RMSD. We thank A. Walker for data tracking and assistance with the manuscript and J. Kunkun for IT help during the assessment. This work was supported in part by the Protein Structure Initiative grant U54 GM074961 (ATCG3D), the NIH Roadmap grant P50 GM073197 (JCMPT), and the Multiscale Modeling Tools for Structural Biology NCCR via grant P41 RR012255.

SUPPLEMENTARY INFORMATION

See online article: [S1 \(box\)](#)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

David Rodríguez and Hugo Gutiérrez-de-Terán
 Fundación Pública Galega de Medicina Xenómica. Complejo Hospitalario Universitario de Santiago. A Choupanna, s/n. Santiago de Compostela (SPAIN).

HOMOLOGY MODELLING of the hA_{21} ADENOSINE RECEPTOR. The target sequence was aligned with the sequences of the three candidate template sequences: human b2 adrenergic receptor (PDB code 2RH1),[1] turkey b1 adrenergic receptor (PDB code 2YTA), [2] and bovine rhodopsin (PDB code 1U19).[3] Multiple sequence alignment was performed by ClustalX,[4] using the PAM250 substitution matrix, with open and elongation/gap penalties of 10 and 0.05, respectively followed manual inspection in order to avoid gaps within helices and proper alignment in loops.

Homology models were generated by Modeller v9.3.[5] Cysteines at positions 77 and 166 were identified as the ones involved in the well-known disulfide bridge between H3 and ECL2 in class A GPCRs, thus the condition was imposed in Modeller. Additionally, we detected a potential disulfide bridge between residues 146 and 159 (ECL2-ECL2). We explored the secondary structure features of ECL2 with three different servers: JSPRED,[6] JPRED,[7] and APSSP2.[8] All of them agreed in a consensus beta-sheet secondary structure between positions 163 to 168 and the corresponding restraint was added. Separate modelling was done considering *i)* the use of one of the three crystallized GPCRs as a template, *ii)* combination of two templates *iii)* imposing or not the potential second disulfide bridge. The program generated 8 models each time, the quality of which was assessed by looking at the stereochemistry with Procheck [9] and visual inspection, considering available mutagenesis data [10]. Selected models were further refined with the aid of the Molprobity server[11].

LIGAND-PROTEIN DOCKING: The ligand ZM241385 was drawn in 2D with Marvin Sketch (Chemaxon, Inc) and a single 3D reliable conformation generated with OMEGA.[12] Automated docking was carried out by GOLD.[13] allowing full flexibility for the ligand, using a sphere of 12 or 15 Å centered on the Cε of HIS278 (necessary radius varied with the protein model used) and 20 GA runs per docking. Both Chemscore and Goldscore scoring functions were considered.[14]

The best complexes were obtained using the protein models generated considering the structure of hβ2 adrenergic receptor (2RH1) as template. 5 complexes from 3 different models were selected and energy minimized with MOE, using MMFF94x force field. Final solutions were ranked attending to the orientation of side chains of the residues that were reported to be important by mutagenesis data [10] in ZM241385 antagonism.

REFERENCES

1. Cherezov et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. *Science* **318**, 1258-1265 (2007).
2. Warne et al. Structure of a beta1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486-491 (2008).
3. Okada et al. The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* **342**, 571-583 (2004).
4. Larkin M.A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948 (2007).
5. Salt, A. & Blindell, T.L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815 (1993).
6. Bryson, K. et al. Protein structure prediction servers at University College London. *Nucl. Acids Res.* **33**, W36-38 (2005).
7. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucl. Acids Res.* **36**, W197-W201 (2008).
8. Raghava, G. P. S. APSSP2: A combination method for protein secondary structure prediction based on neural network and example based learning. *CASP5*, A-132 (2002)
9. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. Appl. Cryst.* **26**, 283 (1993)
10. Martinelli, A. & Tuccinardi, T., Molecular Modeling of Adenosine Receptors: New Results and Trends. *Med Res Rev* **28**, 247-277 (2008).
11. Davis, I. W. et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucl. Acids Res* **35**, W375-W383 (2007).
12. Kirchmair, J., Wolber, G., Lagner, C. & Langger, T. Comparative performance assessment of the conformational model generators omega and catalyst: a large-scale survey on the retrieval of protein-bound ligand conformations. *J. Chem. Inf. Model.* **46**, 1848-61 (2006).
13. Jones, G., Willott, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.*, **267**, 727-748, (1997)
14. Verdouk, M. L., Cole, J. C., Hartshorn, M. J., Murray C. W. & Taylor, R. D. Improved Protein-Ligand Docking Using GOLD. *Proteins*, **52**, 609-623, (2003)

Paper III

Yaziji V., **Rodríguez D.**, Gutiérrez-de-Terán H., Coelho A., Caamaño O.,
García-Mera X., Cadavid M.I., Brea J., Loza M.I. and Sotelo E.

Pyrimidine Derivatives as Potent and Selective A₃ Adenosine Receptor
Antagonists

J. Med. Chem., **2011** 54(2):457-71

Pyrimidine Derivatives as Potent and Selective A₃ Adenosine Receptor Antagonists

Vicente Yaziji,^{†,‡} David Rodríguez,[§] Hugo Gutiérrez-de-Terán,^{*,§} Alberto Coelho,^{†,‡} Olga Caamaño,[#] Xerardo García-Mera,[#] José Brea,^{||} María Isabel Loza,^{*,||} María Isabel Cadavid,^{||} and Eddy Sotelo^{*,†,‡,#}

[†]Combinatorial Chemistry Unit (COMBIOMED), Institute of Industrial Pharmacy, [‡]Center for Research in Biological Chemistry and Molecular Materials, University of Santiago de Compostela, [§]Public Galician Foundation of Genomic Medicine, Santiago de Compostela University Hospital (CHUS), 15706, Santiago de Compostela, Spain, [#]Department of Organic Chemistry, and ^{||}Department of Pharmacology, Faculty of Pharmacy, University of Santiago de Compostela, Santiago de Compostela, 15782, Spain

Received June 11, 2010

Two regioisomeric series of diaryl 2- or 4-amidopyrimidines have been synthesized and their adenosine receptor affinities were determined in radioligand binding assays at the four human adenosine receptors (hARs). Some of the ligands prepared herein exhibit remarkable affinities ($K_i < 10$ nM) and, most noticeably, the absence of activity at the A₁, A_{2A}, and A_{2B} receptors. The structural determinants that support the affinity and selectivity profiles of the series were highlighted through an integrated computational approach, combining a 3D-QSAR model built on the second generation of GRIND INdependent Descriptors (GRIND2) with a novel homology model of the hA₃ receptor. The robustness of the computational model was subsequently evaluated by the design of new derivatives exploring the alkyl substituent of the exocyclic amide group. The synthesis and evaluation of the novel compounds validated the predictive power of the model, exhibiting excellent agreement between predicted and experimental activities.

Introduction

The ubiquitous nucleoside adenosine is essential for the proper functioning of every cell in mammalian species. Adenosine is directly linked to energy metabolism through ATP, ADP, and AMP, while at the extracellular level it regulates a wide range of biological functions through activation of specific receptors (adenosine receptors, ARs),^{1–3} which are classified as A₁, A_{2A}, A_{2B}, and A₃⁴ and belong to the superfamily of the G-protein coupled receptors (GPCRs^o). The improved understanding of the physiology, pharmacology, structure, and molecular biology of adenosine and its receptors has provided solid foundations that support the potential the development of conceptually unexplored therapeutic strategies to address serious unmet medical needs. The advances in the medicinal chemistry of this emerging family of therapeutics have been reviewed recently.^{5–7}

The A₃AR subtype is the most recently characterized member of the family.⁸ Activation^{9,10} of this subtype has been shown to inhibit adenylate cyclase, to increase phosphatidylinositol-specific phospholipase C and D activity, to elevate intracellular Ca²⁺ and IP₃ levels, and to enhance the release of inflammatory and allergic mediators from mast cells. The

therapeutic applications derived from the modulation of this receptor subtype have been reviewed recently.^{11–18} In particular, it is becoming increasingly apparent that antagonists of A₃AR might be therapeutically useful for the acute treatment of stroke and glaucoma,¹⁹ inflammation,^{20–22} and in the development of cerebroprotective,^{23,24} antiasthmatic and antiallergic drugs.^{25,26} Furthermore, recent evidence^{27–31} of high levels of expression of A₃ARs in several cell lines has suggested potential applications for A₃AR antagonists in cancer chemotherapy.

The putative applications of these compounds as drugs, as well as the growing demand for pharmacological tools to study the human A₃AR roles, has made the identification of potent and selective small molecule antagonists of this receptor subtype a topic of great interest.^{11–18} The search for A₃AR antagonists began with the observation that xanthines — a successful structural motif in the search for antagonists for the other ARs subtypes — exhibit low binding affinities for the A₃ receptor subtype. The pursuit of A₃AR antagonists therefore focused on the exploration of structurally diverse heterocyclic libraries. Nowadays, the best known class of A₃AR ligands (Figure 1) includes highly diverse families of tri- and bicyclic heteroaromatic scaffolds and, to a lesser extent, mono-heterocyclic systems. Whereas the systematic structural elaboration of these prototypes has provided derivatives possessing good affinity,^{11–18} the selectivity issue and the relatively poor bioavailability profiles of drug candidates have remained elusive until recently.^{5–7}

The pyrimidine core, being part of the heterocyclic moiety of the endogenous ligand of these receptors (adenosine), is a recurrent substructural motif within bi- and tricyclic ARs antagonists.^{5–7} The well-documented contributions of this chemotype to the field notwithstanding, relatively few papers

*To whom correspondence should be addressed. (H.G.) tel.: +34-881813873, e-mail: hugo.teran@usc.es. (M.L.L.) tel.: +34-881815005, e-mail: mabel.loza@usc.es. (E.S.) tel.: +34-881815221, Fax: +34-981-528093, e-mail: eddy.sotelo@usc.es.

^o Abbreviations: hARs, human adenosine receptors; ADP, adenosine diphosphate; ATP, adenosine-5'-triphosphate; AMP, adenosine monophosphate; GPCRs, G-protein coupled receptors; 3D-QSAR, three-dimensional quantitative structure–activity relationships, GRIND2, GRIND INdependent Descriptors; IP₃, inositol trisphosphate; CHO cells, Chinese hamster ovary cells; CLACC, consistency large auto and cross correlation; MIF, molecular interaction fields; PDB, Protein Data Bank; LOO, leave-one out; rmsd, root mean square deviation.

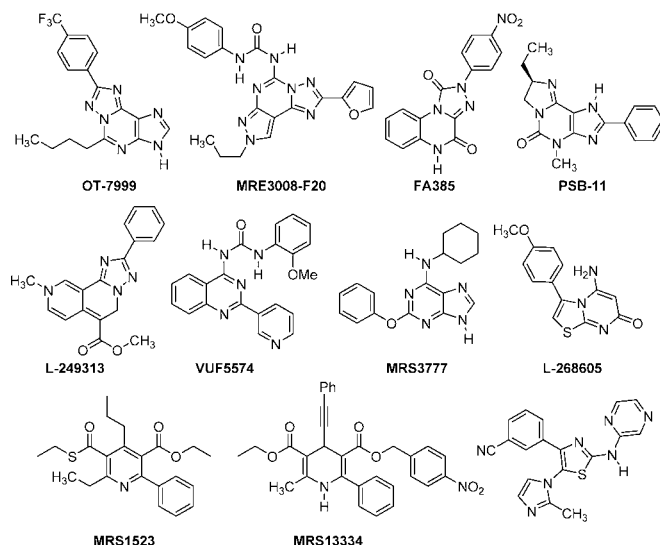
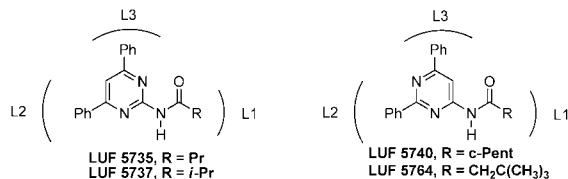


Figure 1. Structures of representative selective A_3 adenosine receptor antagonists.



Code	hA_1^V	hA_{2A}^V	hA_{2B}^V	hA_3^V	Code	hA_1^V	hA_{2A}^V	hA_{2B}^V	hA_3^V
LUF5735	3.70	7%	54%	38%	LUF5740	2.14	196.00	52%	170.00
LUF5737	8.87	44%	79%	45%	LUF5764	8.75	32%	4%	39%

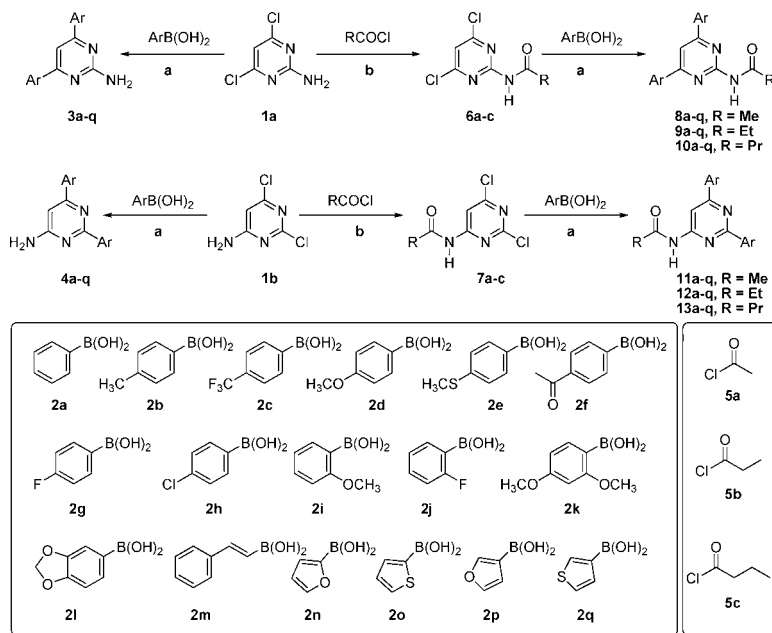
^V Adenosine binding data is expressed as K_i (nM) or as % of displacement at $1 \mu M$.³⁹

Figure 2. Structures and biological data for representative diphenyl 2- or 4-amidopyrimidines as selective A_1 adenosine receptor antagonists. The substitutions further explored in the present report follow those established in the early A_1 AR model,³⁹ labeled as L1, L2, and L3, indicating the lipophilic pockets in the receptor. Note that there is no substituent in L1 in the series of aminopyrimidines and that L2=L3 for all compounds described herein.

have concerned focused programs based on this scaffold,^{32–36} or its biososters (e.g., triazines).³⁷ Two recent publications have covered (1) a molecular simplification study from triazoloquinoxalines to pyrimidines³⁸ and (2) the elaboration of a pharmacophoric model for A_1 adenosine receptors, based on structurally simple regioisomeric diarylpyrimidine scaffolds (Figure 2).³⁹ The latter work not only enabled the identification of potent and selective A_1 AR antagonists derived from either the 4,6-diphenyl-2-amidopyrimidine or 2,6-diphenyl-4-amidopyrimidine templates (Figure 2) but also provided a valuable structural model that could be exploited for the design of new series of compounds.

In light of these precedents, and particularly residual activity toward the hA_3 AR subtype observed for some previously reported compounds (Figure 2),³⁹ it was envisioned that the structural redecoration of the aryl fragments on the amidopyrimidine templates would modify the adenosine receptor selectivity profile and provide new selective A_3 AR antagonists. We therefore focused on the exploration of

diverse aryl moieties on the heterocyclic scaffold, with particular attention paid to structural elements that had been previously identified as contributors in the molecular recognition of the A_3 AR subtype (e.g., 4-methoxyphenyl group).^{40–43} From methodological and practical points of view, it was decided first to explore the synthesis and screening of libraries incorporating identical aryl groups at positions 4,6 and 2,6, as a proof of concept. Thereafter, depending on the results of this first series (reported in the current manuscript) the synthesis of nonidentical series will be performed. The design of the new chemical entities was assisted, and interpreted, by developing an integrated molecular modeling approach that combined ligand docking and 3D quantitative-structure activity (3D-QSAR) studies. Although limitations in the homology modeling of ARs in the design of new ligands have recently been recognized,^{44–46} the recent release of the crystal structure of human A_{2A} AR in complex with the potent inhibitor 4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-d][1,3,5]triazin-5-ylamino]ethyl)phenol (ZM241385)⁴⁷ has been a breakthrough in this

Scheme 1^a

^a Reagents: (a) Pd[(PPh)₃]₄, DME/H₂O, Na₂CO₃, (b) THF, TEA.

area, as occurred earlier with the release of the structure of the $h\beta_2$ adrenergic receptor.⁴⁹ In fact, the most recent models of the A₃AR have already taken advantage of this crystal structure in the description of receptor–antagonist recognition.⁴⁸ On the other hand, structure-based approaches have frequently been combined in G protein-coupled receptor (GPCR) research with ligand-based techniques, such as pharmacophore models³⁹ or 3D-QSAR studies.⁴⁶ In the present work, a new A₃AR model, derived from the recent crystal structure of A_{2A}AR, is reported and used as a basis for the automated docking of the series reported here. In a first iteration, an initial batch of compounds was synthesized, tested, and computationally investigated for the binding mode of the series, which guided the design of the rest of the compounds series. Once the experimental affinities were available for the compounds here reported, the new ligands were computationally described and their structure–affinity was modeled by using the most recent version of the GRIND INdependent Descriptors (GRIND-2),^{50,51} thus providing a rational interpretation of the structure–activity and structure–selectivity relationships. To further challenge the computational model in terms of robustness and predictive capability, it was used for the design of novel compounds bearing new alkyl substitutions on the L1 site (see Figure 2). The synthesis and evaluation of the novel compounds validated the predictive power of the model, exhibiting excellent agreement between predicted and experimental activities.

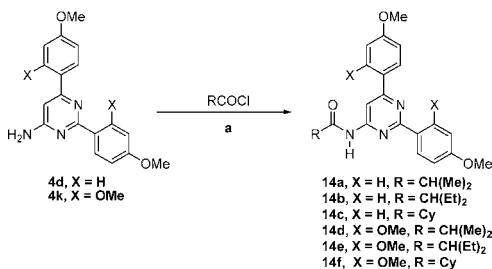
Chemistry

Given that the feasibility of the proposed aim is heavily reliant on the exhaustive exploration of diverse (hetero)aryl

residues on the functionalized pyrimidine templates, a short and divergent synthetic strategy was optimized.⁵² The synthetic pathway developed to access the designed regioisomeric libraries is presented in Scheme 1, and this relied on the commercial availability of the 2- or 4-aminodichloropyrimidines **1a–b** as precursors. Application of the standard conditions of the highly reliable and well-established Suzuki–Miyaura cross-coupling reaction to a collection of commercially available boronic acids (**2a–q**), which representatively cover both the aryl and heteroaryl series (Scheme 1), enabled the rapid decoration of the heterocyclic core at positions 2,4- or 4,6- to afford diarylpyrimidinamines **3a–q** and **4a–q**, which can be considered biosisters of previously described 2-amino-4,6-diaryltriazines.³⁷ Derivatization of the amine function in the heterocyclic precursors **1a–b** by treatment with acid chlorides **5a–c** and subsequent palladium-catalyzed (hetero)arylation afforded two regioisomeric series of di(hetero)aryl 2- or 4-amidopyrimidines (**8–13**).

In an attempt to validate the robustness and predictive capability of the herein developed computational model some computer-generated new ligands, designed to evaluate the tolerance of A₃AR to the introduction of bulky alkyl residues in the amide moiety (L1) of the pyrimidin-4-amine series, were prepared. Treatment of two representative amines, incorporating binding residues that conferred high A₃AR affinity (**4d** and **4k**), with three additional acid chlorides (Scheme 2) afforded the new structures **14a–f**.

The synthetic program provided a focused library of 142 members, which in turn can be subdivided into two regioisomeric sublibraries [34 amines (**3–4**) and 108 amides (**8–14**)] that were structurally characterized. A detailed account of the experimental procedures and the complete description of the

Scheme 2^a

^a Reagents: (a) THF, TEA.

analytical and spectroscopic data for all compounds are available in the Supporting Information.

Biological Evaluation

The affinities of the obtained compounds at the four human adenosine receptor subtypes were determined *in vitro* using radioligand binding assays according to experimental protocols described elsewhere.⁵³ Human adenosine receptors expressed in transfected CHO (A₁AR), HeLa (A_{2A}AR and A₃AR), and HEK-293 (A_{2B}AR) cells were employed. (³H)-1,3-Dipropyl-8-cyclopentylxanthine ((³H)DPCPX) for A₁AR and A_{2B}AR, [³H]4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol for A_{2A}AR, and [³H]NECA for A₃AR were employed as radioligands in binding assays. The biological data (Tables 1–3) are expressed as $K_i \pm$ SEM (nM, $n = 3$) or percentage of inhibition of specific binding at 0.1 μ M ($n = 2$, average) for those compounds that did not fully displace radioligand binding.

Functional Assay at Adenosine A₃ Receptors

Some representative ligands that show affinity toward the hA₃AR subtype were also studied through cAMP experiments (see Figure 3 and Table 3). The functional evaluation was carried out with intact cells expressing the hA₃AR. The inhibition of forskolin-stimulated cAMP production by the receptor agonist was used as a read-out. Concentration–response curves of two representative compounds (compounds **11d** and **12d**) over 0.1 μ M NECA-induced A₃AR activation are shown in Figure 3. cAMP formation was measured by enzyme immunoassay (GE Healthcare). Antagonistic potency, measured as K_B , was calculated from the formula: $K_B = (IC_{50}) / ((2 + ([A] / [A_{50}]^n)^{1/n}) - 1)$, where IC_{50} is the concentration of the antagonist that inhibits the agonist stimulation by 50%, $[A]$ is the concentration of the agonist in the assay, $[A_{50}]$ is the concentration of the agonist that elicits the half-maximum response, and n is the slope of the concentration response curve.⁵⁴

All these derivatives fully reverted the A₃AR-elicited inhibition of cAMP accumulation, unequivocally validating the antagonistic behavior of these compounds at the human A₃AR. Moreover, a comparative analysis of the K_B values for these compounds (Table 3) during the cAMP experiments revealed a clear correspondence with the affinity values determined during the binding experiments (K_i in Table 4).

Structure–Activity Relationship and Molecular Modeling

Affinities in radioligand binding assays at the four human adenosine receptors (A₁, A_{2A}, A_{2B}, and A₃) are reported for

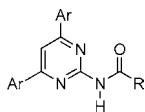
the 4,6-diaryl-2-amidopyrimidines (**8a–q–10a–q**, Table 1) and the 2,6-diaryl-4-amidopyrimidines (**11a–q–13a–q**, Table 2), as well as for the isomeric amine series (**3a–q** and **4a–q**, see Supporting Information). Examination of the binding data indicates that new potent and highly selective ligands for the A₃ receptor subtype have been identified (Table 2, compounds **11b**, **11d**, **12d**, **13d**, **11f**, **12h**, **11j**, **11k**, **12k**, **13k**, **11m**). These results validate the initial hypothesis that the appropriate decoration of the heterocyclic scaffold with previously unexplored diversities would lead to remarkable modifications in the pharmacological activity in comparison to the published results for analogous compounds. Moreover, the documented data exemplify how the structural manipulation of these privileged scaffolds is able to modify the biological profile, not only at the quantitative (affinity) level but also at the qualitative (selectivity) level.

Bearing in mind the considerable number of compounds tested, and for the sake of brevity and clarity, the analysis and interpretation of the data will be carried out at two levels. On the one hand, the most prominent features of the structure–activity (SAR) and structure–selectivity (SSR) relationships for both series will be discussed qualitatively. On the other hand, a more in-depth and quantitative structure–activity relationship can be obtained on the basis of an integrated molecular modeling study. Such an analysis was performed on the set of 64 compounds with experimental K_i values in the A₃ receptor, and this represents a novel approach based on the combination of molecular docking on a homology model for the A₃AR and a 3D-QSAR study.

It can be observed from the biological data the amine series (**3a–q** and **4a–q**) did not exhibit attractive pharmacological profiles at any of the ARs (see Supporting Information). The moderate affinity toward the A₁ receptor subtype elicited by the parent compounds of the regioisomeric series (Ar = Ph, compounds **3a** and **4a**) was generally extinguished by the introduction of groups at the phenyl rings or their replacement by diverse heterocyclic cores. The generally disappointing binding data are common to both regioisomeric amine subsets (**3** and **4**). The most remarkable derivative within the series (**4l**) combines a potent A₁AR antagonistic effect ($K_i = 7.99$ nM) and a satisfactory selectivity (> 30) versus the human A₃AR subtype.

Inspection of the pharmacological data obtained for the most populated set of compounds prepared in this work (i.e., the diaryl 2- or 4-amidopyrimidines **8a–q–13a–q**, Tables 1 and 2) confirms that the systematic modification of the structural prototypes produced a significant, but differentiated, variation in their biological behavior. A comparative analysis of these data highlights the different activity profiles elicited for the two regioisomeric series toward ARs (Tables 1 and 2). Thus, compounds that incorporate the amide moiety at position 4 of the heterocyclic core afforded the most interesting derivatives identified during this study, while their regioisomeric congeners gave a somewhat poor activity profile. Within the 2-amidopyrimidine series only those ligands bearing tolyl groups at positions 4 and 6 of the heterocyclic core (compounds **8b**, **9b**, and **10b**) and the *N*-[2,6-di(benzod[1,3]dioxol-5-yl)pyrimidin-4-yl]acetamide (**8l**) elicited moderate A₃AR affinity (Table 1).

In clear contrast to previously discussed results for the amines (**3** and **4**) and 4,6-diaryl-2-amidopyrimidines (**8–10**), the biological data obtained for the 2,6-diaryl-4-amidopyrimidine subset (Table 2, compounds **11–13**) unequivocally show the determinant influence that the varied structural

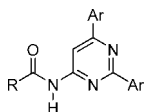
Table 1. Structure and Affinity Binding Data for the 4,6-Diaryl-2-amidopyrimidines **8**, **9**, and **10** at the Human Adenosine Receptors

comp	Ar	R	K_i (nM) or % at 0.1 μ M			
			hA_{1i}^a	hA_{2A}^b	hA_{2B}^c	hA_{3i}^d
8a	Ph	Me	17%	11%	6%	13%
9a		Et	31.3 \pm 2.1	554 \pm 32	42.5 \pm 3.4	531 \pm 31
10a		Pr	10.3 \pm 1.9	3%	14%	17%
8b	4-Me-Ph	Me	3%	6%	3%	47.3 \pm 4.7
9b		Et	1%	1%	3%	157 \pm 22
10b		Pr	3760 \pm 225	2%	10%	131 \pm 19
8c	4-CF ₃ -Ph	Me	2%	1%	2%	10%
9c		Et	2%	12%	9%	2%
10c		Pr	1%	2%	13%	2%
8d	4-MeO-Ph	Me	1%	7%	1%	20%
9d		Et	2%	1%	8%	13%
10d		Pr	2%	24%	9%	15%
8e	4-MeS-Ph	Me	1%	1%	2%	14%
9e		Et	505 \pm 36	183 \pm 16	374 \pm 22	4435 \pm 85
10e		Pr	3%	7%	2%	3%
8f	4-MeCO-Ph	Me	7%	8%	14%	2%
9f		Et	1%	12%	1%	17%
10f		Pr	7%	14%	2%	22%
8g	4-F-Ph	Me	2%	2%	1%	8%
9g		Et	14%	3%	1%	22%
10g		Pr	18%	5%	2%	27%
8h	4-Cl-Ph	Me	2%	2%	1%	1%
9h		Et	1%	11%	2%	14%
10h		Pr	1%	3%	1%	17%
8i	2-F-Ph	Me	8%	12%	10%	15%
9i		Et	48.2 \pm 4.6	424 \pm 38	307 \pm 19	279 \pm 24
10i		Pr	23.5 \pm 1.8	647 \pm 101	6%	15%
8j	2-MeO-Ph	Me	4%	10%	1%	2%
9j		Et	21%	13%	2%	22%
10j		Pr	13%	11%	1%	18%
8k	2,4-MeO-Ph	Me	1%	2%	1%	9%
9k		Et	3%	14%	3%	17%
10k		Pr	14%	16%	15%	1%
8l	3,4-(CH ₂ -O ₂)-Ph	Me	1%	15%	1%	101 \pm 7
9l		Et	535 \pm 37	16%	1%	127 \pm 4
10l		Pr	678 \pm 42	8%	1%	751 \pm 18
8m	Ph-CH=CH-	Me	1%	15%	1%	4%
9m		Et	1%	1%	1%	142 \pm 9
10m		Pr	1%	3%	2%	12%
8n	2-furan	Me	1%	19%	11%	9%
9n		Et	24%	15%	21%	1903 \pm 116
10n		Pr	17%	23%	15%	1%
8o	2-thiophene	Me	2%	5%	3%	21%
9o		Et	24%	25%	2%	21%
10o		Pr	255 \pm 31	201 \pm 28	3%	24%
8p	3-furan	Me	1%	1%	1%	1%
9p		Et	3%	2%	11%	16%
10p		Pr	11%	17%	0%	21%
8q	3-thiophene	Me	2%	2%	1%	1%
9q		Et	367 \pm 25	1893 \pm 174	3250 \pm 261	1279 \pm 215
10q		Pr	85.7 \pm 5.2	4568 \pm 251	12%	2%

^aDisplacement of specific [³H]DPCPX binding in human CHO cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^bDisplacement of specific [³H]4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^cDisplacement of specific [³H]DPCPX binding in human HEK-293 cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^dDisplacement of specific [³H]NECA binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$).

parameters have on the antagonistic profile of these series. The exhaustive exploration of the scaffold enabled the

identification of structurally simple derivatives that exhibit outstanding affinity and remarkable selectivity for the A₃AR

Table 2. Structure and Affinity Binding Data for the 2,6-Diaryl-4-amidopyrimidines **11**, **12**, and **13** at the Human Adenosine Receptors

comp	Ar	R	K_i (nM) or % at 0.1 μ M			
			hA_1^a	hA_{2A}^b	hA_{2B}^c	hA_3^d
11a	Ph	Me	31.2 \pm 4.1	255.3 \pm 13	19%	12.1 \pm 1.3
12a		Et	22.3 \pm 3.3	84.5 \pm 5.7	76.6 \pm 6.4	45.5 \pm 7.4
13a		Pr	19.5 \pm 3.2	103 \pm 8	1%	171 \pm 21
11b (ISVY133)	4-Me-Ph	Me	2%	8%	2%	4.4 \pm 0.3
12b		Et	36.9 \pm 4.1	1%	2%	18.3 \pm 1.9
13b		Pr	16%	2%	3%	59.0 \pm 2.3
11c	4-CF ₃ -Ph	Me	1%	1%	1%	126 \pm 11
12c		Et	3%	17%	8%	12%
13c		Pr	3%	1%	6%	12%
11d (ISVY130)	4-MeO-Ph	Me	1%	10%	4%	3.6 \pm 0.2
12d (ISVY074)		Et	8%	1%	3%	3.6 \pm 0.40
13d (ISVY071)		Pr	8%	4%	1%	11.0 \pm 1.3
11e	4-MeS-Ph	Me	1%	1%	2%	71.3 \pm 3.5
12e		Et	1%	1%	3%	43.6 \pm 1.7
13e		Pr	3%	2%	1%	12%
11f	4-MeCO-Ph	Me	10%	13%	16%	25.2 \pm 0.7
12f		Et	1%	1%	2%	43.9 \pm 2.4
13f		Pr	2%	1%	1%	133 \pm 20
11g	4-F-Ph	Me	15%	1334 \pm 110	1%	16.7 \pm 1.4
12g		Et	83.9 \pm 5.0	429 \pm 18	1%	12.1 \pm 0.6
13g		Pr	82.3 \pm 3.4	1829 \pm 47	2%	34.8 \pm 3.1
11h	4-Cl-Ph	Me	1%	10%	1%	63.3 \pm 8.2
12h		Et	16%	20%	3%	25.3 \pm 0.5
13h		Pr	16%	21%	1%	103 \pm 6
11i	2-F-Ph	Me	17%	73.8 \pm 6.0	21%	18.1 \pm 0.7
12i		Et	31.6 \pm 4.1	103 \pm 5	16%	160 \pm 14
13i		Pr	18.7 \pm 2.5	142 \pm 7	9%	135 \pm 11
11j	2-MeO-Ph	Me	1%	14%	1%	24.1 \pm 1.3
12j		Et	113 \pm 9	22%	7%	23.2 \pm 0.8
13j		Pr	41.3 \pm 2.6	14%	3%	110 \pm 15
11k (ISVY167)	2,4-MeO-Ph	Me	1%	6%	2%	5.4 \pm 0.1
12k (ISVY169)		Et	6%	14%	8%	11.3 \pm 1.4
13k		Pr	18%	12%	14%	10.2 \pm 1.1
11l	3,4-(CH ₂ -O ₂)-Ph	Me	17.7 \pm 3.1	3345 \pm 127	2%	3.3 \pm 0.3
12l		Et	5.28 \pm 0.8	2541 \pm 64	1668 \pm 39	14.5 \pm 1.2
13l		Pr	9.7 \pm 1.4	22%	16%	59.0 \pm 4.3
11m	Ph-CH=CH-	Me	1%	13%	1%	15.6 \pm 2.1
12m		Et	17%	2%	2%	46.9 \pm 5.4
13m		Pr	1%	1%	8%	25%
11n	2-furan	Me	40.7 \pm 5.2	8.1 \pm 1.2	12.0 \pm 1.1	3.0 \pm 0.4
12n		Et	15.5 \pm 3.1	6.4 \pm 0.7	20.5 \pm 2.4	6.2 \pm 0.7
13n		Pr	7.8 \pm 0.9	5.7 \pm 0.4	16.4 \pm 0.7	9.9 \pm 1.2
11o	2-thiophene	Me	19%	24.6 \pm 2.6	23%	8.0 \pm 0.4
12o		Et	32.9 \pm 1.4	114 \pm 7	17%	21.8 \pm 2.2
13o		Pr	33.3 \pm 3.2	153 \pm 11	8%	23.0 \pm 4.0
11p	3-furan	Me	1%	74.0 \pm 3.2	302 \pm 67	10.1 \pm 0.9
12p		Et	132 \pm 10	82.8 \pm 6.0	49.0 \pm 5.3	12.6 \pm 1.1
13p		Pr	65.1 \pm 4.0	544 \pm 11	1%	3%
11q	3-thiophene	Me	13%	39.8 \pm 4.4	16%	20.2 \pm 3.2
12q		Et	39.4 \pm 2.3	63.5 \pm 4.1	164.7 \pm 47	64.2 \pm 5.6
13q		Pr	65.1 \pm 4.7	23%	21%	11%

^aDisplacement of specific [³H]DPCPX binding in human CHO cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^bDisplacement of specific [³H]4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylaminoethyl)phenol binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^cDisplacement of specific [³H]DPCPX binding in human HEK-293 cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^dDisplacement of specific [³H]NECA binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$).

(see Table 2, compounds **11b**, **11d**, **12d**, **13d**, **11k**, **12k**, **13k**). A comparison of these data with the observed activity for the

parent compounds of the series (Table 2, Ar = Ph, compounds **11a**, **12a**, **13a**, relatively A₁AR potent but somewhat

Table 3. Antagonistic Potency (Measured as K_B) at Human A_3 Receptors of Selected Compounds^a

compound	K_B (nM)
12d	1.40 ± 0.09
13d	1.57 ± 0.13
11k	3.85 ± 0.41
11b	3.56 ± 0.23
12h	3.88 ± 1.17

^a Values represent the mean ± SEM of two separate experiments.

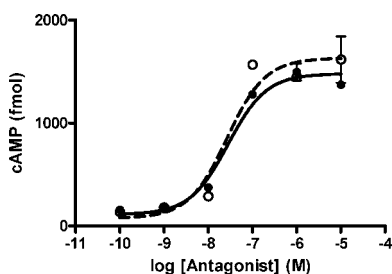


Figure 3. Effect of **11d** (○, dashed fitting) and **12d** (●, black fitting) on 0.1 μM NECA-induced cAMP decrease of 10 μM forskolin-stimulated human A_3 receptors. Points represent the mean ± SEM (vertical bars) of two separate experiments.

promiscuous AR ligands) allows the rapid evaluation of the effects caused by the structural modifications. In general, modification of the aromatic substitution pattern completely extinguished the affinity for the A_1 AR, while conferring notable potency and selectivity toward the A_3 AR subtype. Remarkably, such a subtle structural modification is able to produce a radical variation in the activity profile, being significant not only for a methoxy group at position 4 but also for the more highly diverse residues explored (e.g., methyl, thiomethyl, acetyl, fluoro and chloro). It is also remarkable that the vinyl analogues (Table 2, compounds **11m** and **12m**) of the parent compounds proved to be relatively potent and highly selective A_3 ligands, a finding that reaffirms how bulky substituents at sites L2/L3 favor selectivity toward hA_3 AR. The consequences of introducing a group at position 2 of the phenyl ring was also briefly assessed (Table 2, compounds **11i**, **12i**, **13i**, **11j**, **12j**, and **13j**). As observed, the introduction of fluoro or methoxy groups at this position afforded relatively potent derivatives, albeit with markedly different selectivity profiles. Within this ligand subset only the 2-methoxyphenyl derivative of the 4-acetamide series (compound **11i**) elicited a satisfactory affinity/selectivity profile. Conversely, the simultaneous introduction of methoxy groups at positions 2 and 4 of the phenyl ring afforded highly potent and completely selective ligands (**11k**, **12k**, and **13k**) toward the A_3 AR subtype, regardless of the alkylic residue present in the amide group at position 4 of the heterocyclic backbone. Finally, in a clear contrast with the results described so far, replacement of the phenyl group in the parent compounds by heterocyclic cores proved to be highly discouraging, generating a series of potent but nonselective ligands.

An integrated analysis of the data presented in Table 2 for the 4-amide homologous series (compounds **11–13**) is shown in Figure 4. In this representation the experimental K_i values at hA_3 AR are plotted as a function of both the L1 and the

L2/L3 substitutions. The weak modulator effect exerted by the alkyl residues of the amide functions (L1) on the activity/selectivity profile within these series can be observed. It can be clearly appreciated that the size of L1 is inversely correlated with the affinity within each subseries, an observation that is consistent with previous findings.³⁸ Moreover, a detailed inspection of the pharmacological data reported for these series (Table 2) shows that A_3 AR selectivity also increases on reducing the size of the L1 substituent. Only some combinations of L2/L3 substituents show little sensitivity to the nature of the L1 substituent, in particular, compounds incorporating 4-methoxyphenyl (**11d**, **12d**, **13d**) or 2,4-dimethoxyphenyl (**11k**, **12k**, **13k**) residues (i.e., the substituent present in compounds eliciting the highest affinity).

Once the initial hypothesis that hA_3 AR could be more tolerant to bulky L2/L3 substituents had been validated by the SAR data, an exhaustive molecular modeling study was developed to gain new insights into the structure–affinity relationship for the hA_3 AR. A homology model of the hA_3 AR receptor was built using the recently crystallized hA_{2A} AR structure as a template. This model served as a basis for an automated docking exploration of the 64 compounds for which experimental K_i values at the hA_3 AR are reported. The choice of the docking algorithm (GOLD program in combination with the Chemscore scoring function)⁵⁵ is the result of an internal validation of different docking alternatives in order to reproduce the experimental binding pose of 4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol/ hA_{2A} AR (data not shown), a validation that is in agreement with a recent comparative study of ligand docking tools in ARs.⁴⁹ The systematic docking exploration identified one conserved binding mode for both regioisomeric diaryl amidopyrimidine series reported here. This binding mode was found in 62 of the 64 compounds (97% of the cases), and this mode was the top scored pose by Chemscore in 66% of the cases. Moreover, in 63% of the cases, this binding pose was the most populated according to an rmsd tolerance of 1 Å for the clustering. This binding mode is represented in Figure 5 for compound **11d**.

The main anchoring point is a double hydrogen bond of the exocyclic amino/amido group (donating) and its closest nitrogen atom in the pyrimidine ring (N3, accepting) with Asn 6.55 (note the Ballesteros–Weinstein residue numbering⁵⁶), a totally conserved residue of the adenosine receptor family. At the same time, the pyrimidine ring is flanked by the side chain of Phe 5.29, in the second extracellular loop (EL2), and Leu 6.51 in helix 6. This interaction pattern of the aminopyrimidine moiety (π -stacking with Phe 5.29, hydrophobic interactions with Leu 6.51 and hydrogen bonding to Asn 6.55) resembles the experimentally observed binding mode of 4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol with the hA_{2A} AR.⁴⁷ Accordingly, the important role in ligand binding of residues 5.29 and 6.51 has recently been validated in a site-directed mutagenesis study of the A_{2A} AR.⁵⁷ Interestingly, Phe 5.29 is totally conserved in the ARs family, while Leu 6.51 is substituted by a smaller valine in the low-affinity hA_{2B} AR, and the replacement of this residue by an alanine in hA_{2B} AR completely abolishes ligand binding.⁵⁷ As far as position 6.55 is concerned, there is biochemical evidence that suggests the important role of this residue in ligand binding for several ARs, including the A_3 AR.^{58–60} The molecular alignment of the 62 molecules, obtained by ligand-docking, is shown in Figure 6. It can be appreciated that the volume of the L1 and L2 subsites has been

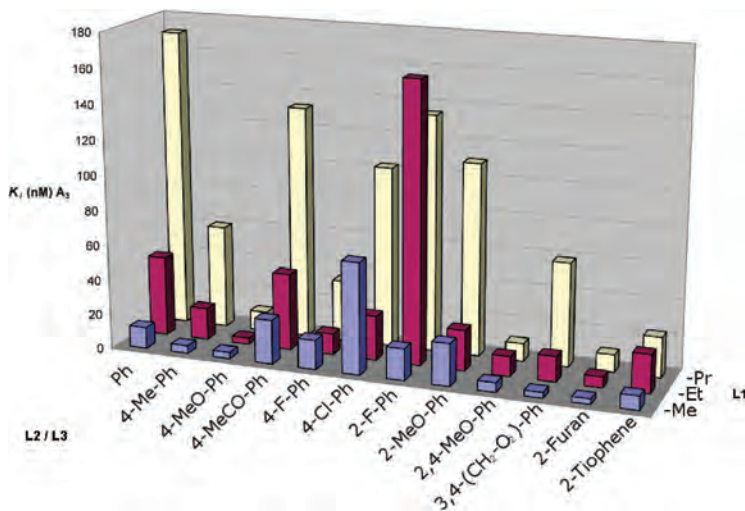


Figure 4. Effect of the nature of the L1 substituent on the human A₃ adenosine receptor affinity for the series of 2,6-diaryl-4-amidopyrimidines. The figure only represents those compounds that have experimental K_i for the three considered L1 substitutions.

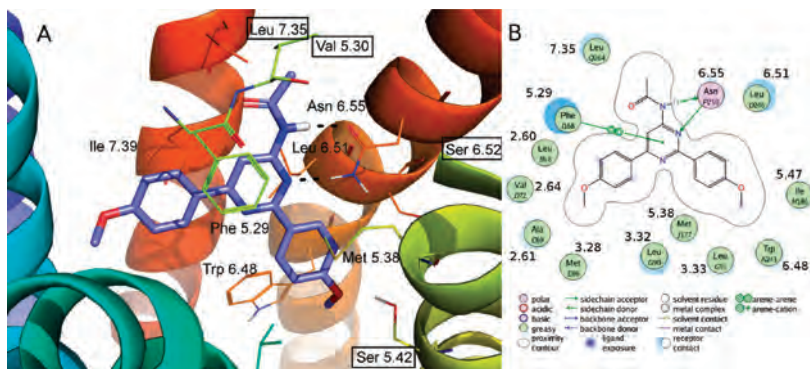


Figure 5. Binding mode of compound **11d**, showing the main receptor–ligand interactions. Residues that are specific for the human A₃ adenosine receptor are shown in boxes in the 3D panel (A), generated in PyMol (<http://www.pymol.org>). The double hydrogen bond with Asn 6.55 is indicated by dashed lines. Panel B shows a schematic representation of residue–ligand interactions, calculated with LigX as implemented in MOE. Residues are labeled according to the Ballesteros & Weinstein numbering.⁵⁶

well explored, while there is a volume tolerance in the subsite occupied by L3 (helices 2, 3, 7) that was not completely explored by our ligand series. Even after one energy minimization cycle, the molecular alignment did not change substantially and the highest variability is still located on the L3 site.

This molecular alignment was the basis for a 3D-QSAR study that involved the use of the new generation of Grid-Independent Descriptors (GRIND-2).^{50,51} The first generation of these molecular interaction field (MIF)-based descriptors was originally conceived precisely to circumvent the necessity of obtaining a highly accurate molecular alignment of the molecules prior to the 3D-QSAR analysis.⁶¹ However, the most recent version of the GRIND methodology includes a new mathematical transformation applied to the MIF descriptors that guarantees that a given variable represents

exactly the same information for every compound of the series.⁵¹ This method, called consistency large auto and cross correlation (CLACC), either generates a molecular alignment of the molecules, on the basis of the correlation of the variables, or either it uses an input molecular alignment provided by the user (e.g., obtained by molecular docking). The first (default) option is recommended for the exploration of compounds that have closely related structures, while the second approach (docking alignment) is a good compromise for series that present problems with the CLACC alignment. An advantage of the last option is that the interpretation of the derived models can be easily expressed in the context of receptor–ligand interactions, allowing it to retrieve structural information on the binding site. We explored all of these different settings for the generation of 3D-QSAR models, the

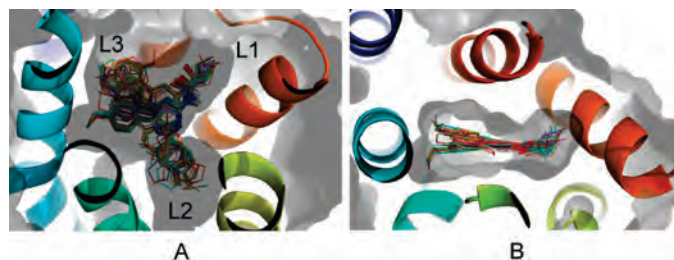


Figure 6. Molecular alignment of the 62 molecules that had the postulated docking pose. This alignment was used as an input for the 3D-QSAR study. (A) side view of the receptor, and (B) view from the extracellular side. The following transmembrane helices (TM) are shown: TM2 (cyan), TM3 (green), TM4 (yellow), TM6 (orange), and TM7 (red). The Connolly surface of the receptor is depicted in gray.

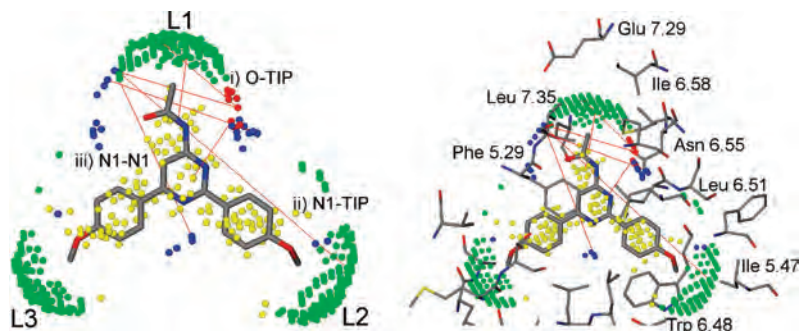


Figure 7. Important variables in the 3D-QSAR model C (represented for compound **11d**). Green dots denote TIP fields, red dots O fields, blue dots N1 fields, and yellow dots DRY fields. (Right) The same representation showing the binding site. The correspondence of the TIP fields with the limiting pockets of the receptor for the L1 (Glu 7.29, Ile 7.35) and L3 (Ile 5.47, Trp 6.48), the superposition of the O–N1 short distance variable with Asn 6.55, and the DRY field with Phe 5.29 can be appreciated.

results being summarized in Table S1 of the Supporting Information. The model generated on the basis of a docking molecular alignment and the CLACC method for encoding the descriptors (*Model C* in Table S1, Supporting Information) was selected for further interpretation. This 3D-QSAR model has two latent variables (LV) and presents a satisfactory statistical quality, with a fitting parameter of $r^2 = 0.86$ and a predictive ability of $q^2 = 0.67$, as obtained by the LOO cross-validation test. The standard error for the correlation and the prediction was 0.31 and 0.48 pK_i log units, respectively.

The interpretation of the model highlights the key structural features for high A_3 AR affinity. The most important variables, that is, those with the highest PLS positive coefficients, were used for the model interpretation and are depicted in Figure 7. These variables represent, in an ideal case, structural features that are present in the active compounds but absent in the inactive compounds. In this respect, the following features are important for the model interpretation:

- (i) The hydrophobic interactions at the extracellular tip of the binding site, mainly with residues Ile 6.58 and Leu 7.35, are identified with the O-TIP (optimum distance at 5.8 Å) and the DRY-TIP (6.6 Å) correlograms. These two hydrophobic residues are a probable source of A_3 specificity: position 6.58 is occupied by a threonine in the other subtypes, while position 7.35 has already been related to interspecies selectivity in the A_1 AR.⁶³ The model suggests that an optimal shape

complementarity is achieved by molecules with smaller L1 substituents (e.g., acetamides) or, alternatively, molecules bearing larger L1 substituents but smaller L2/L3 substituents (e.g., **13n**). The combination of the aforementioned descriptors provides information about the interdependence of the size of L1 and L2/L3 substituents. This descriptor also accounts for the lack of affinity observed in the aminopyrimidine series (see Table 1), since these scaffolds do not bear any alkyl residues on the exocyclic nitrogen.

- (ii) The optimal pharmacophoric distance between the H-bond acceptor probe, corresponding to the carbonyl of the amide, and the shape of the L2 substituent is located at 16.6 Å in the N1-TIP cross-correlogram. Whereas the role of the carbonyl group could be hypothesized as a water-mediated interaction with Glu 7.29 in the third extracellular loop (EL3), this descriptor mainly identifies the importance of residue Ile 5.47, interacting with the L2 substituent. Importantly, Ile 5.47 is occupied by the less bulky valine in the other AR subtypes, a fact that could be taken into account to improve A_3 selectivity.
- (iii) Finally, the O–N1 and N1–N1 autocorrelograms account for the differences between the 2-amido- or 4-amidopyrimidine series, since these molecular descriptors identify the distances between the exocyclic amide and the N1 in the ring. In the series of

	1.35	2.57	2.60	2.61	2.62	2.64	2.65	3.28	3.29	3.32	3.33	3.36	5.28	5.29	5.30	5.35	5.38	5.42	5.43	5.47	6.44	6.46	6.51	6.52	6.54	6.57	6.59	7.34	7.35	7.39	7.42	7.43		
A1	Y	V	L	A	I	I	N	V	A	V	L	T	E	F	M	M	M	N	F	V	F	W	L	H	I	T	H	K	T	Y	I	T	H	
A2A	Y	A	F	A	I	I	S	I	A	V	L	T	L	F	M	M	M	N	F	V	F	W	L	H	I	N	T	H	A	M	Y	I	S	H
A2B	Y	A	F	A	I	I	S	L	A	V	L	T	L	F	E	M	M	N	F	V	F	W	V	H	A	N	T	N	K	M	N	I	S	H
A3	Y	V	L	A	I	V	S	M	T	L	L	T	Q	F	V	M	M	S	F	I	F	W	L	S	I	N	I	E	V	L	Y	I	S	H

Within 2 Å
Within 4,5 Å

Figure 8. Multiple pseudosequence alignment of human ARs, taking hA₃AR residues within 4.5 Å of all docked compounds into account. Residue positions are denoted by the Ballesteros & Weinstein numbering,⁵⁶ and shaded in gray according to their distance toward the docked compounds.

4-amidopyrimidines, optimum distances of 9 and 10.6 Å for O–N1 and N1–N1 autocorrelations, respectively, are observed. Conversely, in the series of 2-amidopyrimidines, a descriptor in the N1–N1 correlation, which is negatively related with affinity in the model, identifies the particular location of the N1 of this scaffold closer to the carbonyl of the amide (distance 4.6 Å).

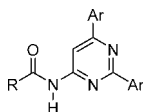
This last point is intriguing, since the docking model does not identify any polar interaction for either the oxygen of the amide group or the nitrogen at position N1 of the pyrimidine (i.e., the nitrogen that varies in position between 2- and 4-amidopyrimidines). However, if we compare the binding mode of the molecular series here reported on the A₃AR with the experimental binding mode of 4-(2-[7-amino-2-(2-furyl)-[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol on the A_{2A}AR, it appears that the N1 in the 4-amidopyrimidines series overlaps with a nitrogen in the heterocycle of the standard (N19 according to PDB nomenclature in entry 3EML).⁴⁷ Recently, the group of Jacobson⁶³ noted the importance of a polar interaction of this N19 with crystallographic water molecules. In order to check if similar interactions could be achieved in our 4-amidopyrimidines/hA₃AR complexes, we performed a computational exploration of structural water molecules in the binding site of the hA₃AR model, as detailed in the methods section. The results (Figure S2, Supporting Information) show an energetically favorable area for a water molecule that overlaps with the N1–N1 descriptor, close to the position of the varying nitrogen in the 4-amidopyrimidines (blue dots at the bottom of Figure 7). A water-mediated interaction between N1 and Thr 7.42, which somehow resembles the H-bond network ZM24-1386(N19)–HOH⁵⁵⁹–HOH⁵⁵⁰–His7.43(Nε) in the A_{2A}AR, is thus proposed as an specific polar contact for the 4-amidopyrimidine series, lacking in the 2-amidopyrimidines. The experimental validation of such a different interaction behavior is currently being investigated in our laboratory by synthesizing additional series of unexplored heterocyclic scaffolds.

As stated above, an advantage of the 3D-QSAR methodology employed in this study is to place the relevant descriptors in the context of the binding site. This procedure can not only deal with the structural requirements of the hA₃AR for high affinity but also enables a comparison of the hot-spots with the relative positions in other ARs in order to explore the reasons for selectivity. In an effort to identify these hot-spots, we built a so-called “pseudosequence” based on the docking results of this study. This pseudosequence is defined by all of the residues of the receptor located at a maximum distance of 4.5 Å from the most exposed atom of the group of ligands docked in the hA₃AR (as superimposed in Figure 6). An

alignment of this pseudosequence for the human members of the ARs family is shown in Figure 8, in which the variable positions are clearly identified.

In this respect, it is remarkable that positions 5.42 and 6.52, at the bottom edge of the binding site, are both occupied by a serine in the hA₃AR (see Figure 5); in the other three human ARs these positions are occupied by Asn 5.42 and His 6.52, respectively. The less voluminous side chain of a serine in these positions would allow the accommodation of bulkier L2 substituents in this subsite at the hA₃AR, thus offering a rationale for the observed receptor selectivity. There are also remarkable differences within the ARs family regarding residues at the top of the binding site: Ile 6.58, interacting with L1, is replaced by a smaller valine in the other ARs. hA₃AR presents a valine at position 5.30 (in the tip of EL2), which replaces a Glu that is conserved in the other three hAR subtypes. According to the A_{2A}AR crystallographic structure, Glu 5.30 hydrogen bonds with the side chain of a His 7.29 in EL3, thus closing the top of the binding site while accepting an additional hydrogen bond from the amino group of 4-(2-[7-amino-2-(2-furyl)-[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol.⁴⁷ The amino derivatives would benefit from this interaction, thus explaining the low selectivity profile displayed by this group of compounds (Table 1). Interestingly, a recent study of a new series of 2-phenylpyrazolo[4,3-*d*]pyrimidin-7-ones already indicated this difference in the flap regions as being responsible for the A₃AR selectivity.⁴⁹ Finally, Leu 7.35 provides a specific hydrophobic subsite for the L1 substituent compared to the more voluminous methionine present at this position in the A_{2A}AR and A_{2B}AR or the polar Thr 7.35 in A₁AR. Importantly, site directed mutagenesis studies have identified this position as being responsible for the interspecies differences in ligand affinities in the A₁ARs.⁶²

To further challenge the computational model in terms of robustness and predictive capability, the tolerance of A₃AR to steric factors imposed by the alkyl residue of the amide function (L1) was explored. Accordingly, six compounds were designed which combined three new bulky residues [e.g., CH(Me)₂, CH(Et)₂, and Cy] on the exocyclic amide group with the scaffold of the amines **4d** and **4k**. The compounds were docked on the A₃AR and queried to the QSAR model, which predicted a good affinity profile for the A₃AR (see Table 3). Figure S2 in the Supporting Information shows how the most bulky compounds (**14c** and **14f**) optimally accommodate the cyclohexyl substituent in the hA₃AR pocket. On the other hand, a superposition with the crystallographic structure of the hA_{2A}AR shows that steric clashes with the L1 substituent of this receptor might occur, as anticipated by the pseudosequence analysis shown in Figure 8. The compounds

Table 4. Structure and Affinity Binding Data for 2,6-Diaryl-4-amidopyrimidines **14** at the Human Adenosine Receptors

comp	Ar	R	K_i (nM) or % at 0.1 μ M				pK_i @ hA_3	
			hA_1^a	hA_{2A}^b	hA_{2B}^c	hA_3^d	exp	pred
14a	4-OMe-Ph	CH(Me) ₂	5%	1%	3%	12.2 ± 0.9	7.91	8.16
14b		CH(Et) ₂	3%	6%	8%	58.1 ± 4.7	7.24	7.82
14c		Cy	5%	1%	1%	32.1 ± 2.4	7.49	7.15
14d	2,4-OMe-Ph	CH(Me) ₂	18%	1%	1%	15.9 ± 1.3	7.8	7.58
14e		CH(Et) ₂	1%	2%	2%	52.4 ± 6.2	7.28	7.57
14f		Cy	1%	1%	4%	56.3 ± 6.1	6.59	6.85

^a Displacement of specific [³H]DPCPX binding in human CHO cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^b Displacement of specific [³H]4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino]ethyl)phenol binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^c Displacement of specific [³H]DPCPX binding in human HEK-293 cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$). ^d Displacement of specific [³H]NECA binding in human HeLa cells expressed as $K_i \pm$ SEM in nM ($n = 3$) or percentage displacement of specific binding at a concentration of 0.1 μ M ($n = 2$).

were then prepared and tested at the four human adenosine receptor subtypes (Table 4). As predicted by the computational exploration described above, the newer derivatives exhibit potent and selective activity profiles, which unequivocally confirms the tolerance of hA_3AR to the size of the L1 substituent. Excellent agreement is found between predicted and experimental affinity values for the six compounds designed and tested in this part of the study [with an impressive low standard error of the prediction (SDEP = 0.37 log pK_i units)], which further confirms the predictive power of the integrated computational model reported in this work, that is, combining a molecular alignment from automated docking with the prediction of activities on the basis of the 3D-QSAR model (see Supporting Information, Figure S3C) was observed. It is worth noting that on the solely basis of the available literature³⁹ data or the herein established SAR (Tables 1 and 2) the synthesis of compounds **14** would not have been advisable. Thus, the modeling exploration enabled us to anticipate attractive activity/selectivity profiles for compounds incorporating hindered fragments at the amide chain (as consequence of the higher tolerability for the L1 subsite of the hA_3AR).

Conclusions

A new series of structurally simple and highly potent ligands that exhibit remarkable selectivity profiles toward the A_3AR has been identified. A previous series of potent and selective A_1AR antagonists was selected, and the subsequent stepwise structural diversification of these model substrates was carried out in order to radically modify the activity/selectivity profiles while simultaneously providing valuable structural information on the requirements for its binding at the hA_3 receptor subtype. Excellent affinity toward the hA_3AR ($K_i \leq 6$ nM) and optimal selectivity profiles ($\leq 10\%$ displacement of 0.1 μ M concentrations at the other ARs) were observed for compounds **ISVY133**, **ISVY130**, **ISVY074**, and **ISVY167**, which incorporate 4-tolyl, 4-methoxyphenyl, and 2,4-dimethoxyphenyl moieties at the 2,6-positions of the heterocyclic backbone. The antagonistic behavior of five representative derivatives of these series was unequivocally validated through functional cAMP experiments. The main

SARs identified were substantiated by an exhaustive molecular modeling study that combined a receptor-driven docking model, which was constructed on the basis of the recently published crystal structure of the $hA_{2A}AR$, and a ligand-based 3D-QSAR model, highlighting the key structural features required for the optimal interaction with the hA_3 receptor subtype in these compounds. The robustness and predictive capabilities of the model were validated by designing novel series of compounds that explore new alkyl residues at the L1 subsite, which show high affinity and selectivity profiles for the hA_3AR . We must note that these compounds would not have been synthesized solely on the basis of the available SAR data on the literature³⁹ or the qualitative SAR established on this work (Tables 1–3). On the contrary, the interest of these compounds was envisioned by the computational modeling exploration, suggesting that the hA_3AR shows higher tolerability for the L1 subsite. Further experiments are currently in progress in our laboratories to prepare new libraries incorporating nonidentical aryl groups at positions 2,6 and 4,6 obtained by adaptation of the herein documented synthetic strategy according to recently published methodologies.^{64–66} The biological profile of these new derivatives will be published in due course.

Experimental Section

Chemistry. Commercially available starting materials, reagents, and solvents were purchased (Sigma-Aldrich) and used without further purification. When necessary, solvents were dried by standard techniques and distilled. After being extracted from aqueous phases, the organic solvents were dried over anhydrous sodium sulfate. The reactions were monitored by thin-layer chromatography (TLC) with 2.5 mm Merck silica gel GF 254 strips, and the purified compounds each showed a single spot; unless stated otherwise, UV light and/or iodine vapor were used for detection of compounds. The Suzuki cross-coupling reactions were performed in coated Kimble vials on a PLS (6 × 4) Organic Synthesizer with orbital stirring. Filtration and washing protocols for supported reagents were performed in a 12-channel vacuum manifold from Aldrich. Purity and identity of all tested compounds were established by a combination of HPLC, elemental analysis, mass spectrometry, and NMR spectra as described below. Purification of isolated products was carried out by column chromatography (Kieselgel 0.040–0.063 mm, E. Merck) or

Table 5. Conditions Used for Radioligand Binding Assays Using A₁, A_{2A}, A_{2B}, and A₃ Human Adenosine Receptors

	A ₁	A _{2A}	2A _{2B}	A ₃
Buffer A	20 mM Hepes, 100 mM NaCl, 10 mM MgCl ₂ , 2 units/mL adenosine deaminase (pH = 7.4)	50 mM Tris-HCl, 1 mM EDTA, 10 mM MgCl ₂ , 2 units/mL adenosine deaminase (pH = 7.4)	50 mM Tris-HCl, 1 mM EDTA, 10 mM MgCl ₂ , 0.1 mM benzamidine, 2 units/mL adenosine deaminase (pH = 6.5)	50 mM Tris-HCl, 1 mM EDTA, 5 mM MgCl ₂ , 2 units/mL adenosine deaminase (pH = 7.4)
Buffer B	20 mM Hepes, 100 mM NaCl, 10 mM MgCl ₂ (pH = 7.4)	50 mM Tris-HCl, 1 mM EDTA, 10 mM MgCl ₂ (pH = 7.4)	50 mM Tris-HCl (pH = 6.5)	50 mM Tris-HCl (pH = 7.4)
plate	GF/C	GF/C	GF/B	GF/B
radioligand	[³ H]DPCPX nM	[³ H]ZM2413853 nM	[³ H]DPCPX35 nM	[³ H]NECA 30 nM
nonspecific binding	10 μM (R)-PIA	50 μM NECA	400 μM NECA	100 μM (R)-PIA
incubation	25 °C/60 min	25 °C/30 min	25 °C/30 min	25 °C/180 min

medium pressure liquid chromatography (MPLC) on a Combi-Flash Companion (Teledyne ISCO) with RediSep prepacked normal-phase silica gel (35–60 μm) columns followed by recrystallization. Melting points were determined on a Gallenkamp melting point apparatus and are uncorrected. The NMR spectra were recorded on Bruker AM300 and XM500 spectrometers. Chemical shifts are given as δ values against tetramethylsilane as internal standard and J values are given in Hz. Mass spectra were obtained on a Varian MAT-711 instrument. High resolution mass spectra were obtained on an Autospec Micromass spectrometer. Analytical HPLC was performed on a Agilent 1100 system using an Agilent Zorbax SB-Phenyl, 2.1 mm × 150 mm, 5 μm column with gradient elution using the mobile phases (A) H₂O containing 0.1% CF₃COOH and (B) MeCN and a flow rate of 1 mL/min. Elemental analyses were performed on a Perkin-Elmer 240B apparatus at the Micro-analysis Service of the University of Santiago de Compostela, the elemental composition of the new compounds agreed to within ±0.4% of the calculated value. The purity of all tested compounds was determined to be >95%. A detailed description of synthetic methodologies as well as analytical and spectroscopic data for all described compounds is included in the Supporting Information.

Pharmacology. Radioligand binding competition assays were performed in vitro as previously described⁵⁵ using A₁, A_{2A}, A_{2B}, and A₃ human adenosine receptors expressed in transfected CHO (A₁AR), HeLa (A_{2A}AR and A₃AR), and HEK-293 (A_{2B}AR) cells. The experimental conditions used are summarized in Table 5. In each instance, aliquots of membranes (15 μg for A₁, 10 μg for A_{2A}AR, 18 μg for A_{2B}AR, and 90 μg for A₃AR) in buffer A (see Table 5) were incubated for the specified period at 25 °C with the radioligand (2–35 nM) and six different concentrations (ranging from 0.1 nM to 1 μM) of the test molecule in a final volume of 200 μL. The binding reaction was stopped by rapid filtration in a multiscreen manifold system (Milipore Ibérica, Madrid, Spain). Unbound radioligand was removed by washing four times with 250 μL of ice-cold buffer B for A₁ and A_{2A} receptors, and six times with 250 μL of ice-cold buffer B for A_{2B}AR and A₃AR (see Table 5). Nonspecific binding was determined using a 50 or 400 μM NECA solution for A_{2A}AR and A_{2B}AR and 10 or 100 μM R-PIA solution for A₁AR and A₃AR, respectively. Radioactivity retained on filters was determined by liquid scintillation counting using Univisol (ICN Biochemicals, Inc.). The binding affinities were determined using [³H]-DPCPX (130 Ci/mmol; GE-Healthcare, Barcelona, Spain) as the radioligand for A₁AR and A_{2B}AR, [³H]4-(2-[7-amino-2-(2-furyl)[1,2,4]triazolo[2,3-*a*][1,3,5]triazin-5-ylamino)ethylphenol (21 Ci/mmol; Tocris, Madrid, Spain) for A_{2A}AR and [³H]-NECA (15.3 Ci/mmol; NEN-Perkin-Elmer Life Sciences, Madrid, Spain) for A₃AR.

The inhibition constant (*K_i*) of each compound was calculated by the expression: $K_i = IC_{50} / (1 + (C/K_D))$; where *IC*₅₀ is the concentration of compound that displaces the binding of radioligand by 50%, *C* is the free concentration of radioligand, and *K_D* is the apparent dissociation constant of each radioligand.

The percentage of displacement of specific binding was calculated by the expression: % of displacement = ((BT – dpm) * 100) / (BT – NSB); where BT is the total binding of the radioligand in the assay, NSB is the nonspecific binding of the radioligand in the assay, and dpm are the radioactive measurements obtained by competing the radioligand binding with a given concentration of the test compound. Unless otherwise specified, results shown in the text and tables are expressed as means ± SEM. Significant differences between two means (*p* < 0.05 or *p* < 0.01) were determined by one-way analysis of variance (ANOVA) and/or by Student's *t* test for nonpaired data.

Molecular Modeling

Model Building. A homology model of the hA₃AR was built using the recently crystallized hA_{2A}AR as a template. The modeling protocol is adapted from our participation in the GPCR Dock 2008 competition.⁴⁴ Briefly, a sequence alignment between the two receptors with Clustal (PAM250 substitution matrix, with open and elongation gap penalties of 10 and 0.05)⁶⁷ was provided to Modeller v9.4.⁶⁸ Fifteen initial models were obtained using standard parameters. In a first stage, the best five models were selected on the basis of Procheck⁶⁹ geometrical quality and DOPE scoring, and these were subjected to geometrical improvement by the Molprobit server.⁷⁰ In a second stage, the best model from the previous step was subjected to loop optimization with the LoopModel routine in Modeler,⁷¹ again generating 15 refined models. We selected the best model on the basis of a compromise between Procheck stereochemical quality and the DOPE energetical ranking. The geometry of the loops in the selected model was refined by partial energy minimization (i.e., nonloop residues were frozen) using the Polak–Ribiere algorithm (convergence criteria 0.05 kcal/mol·Å²) and the OPLS-AA force-field as implemented in MacroModel.⁷² The general numbering scheme for GPCRs proposed by Ballesteros and Weinstein⁵⁶ was adopted through this work. In essence, every residue is numbered as X.YY, where X corresponds to the transmembrane helix (X = [1,7]) and YY is a correlative number in the protein sequence, but taking as a reference position (YY = 50) the most conserved residue in the given helix.

Protein–Ligand Docking. Automated docking exploration was performed with GOLD version 3.2.⁷³ Each ligand was docked 20 times with default (high accuracy) genetic algorithm (GA) search parameters, using the scoring function Chemscore as implemented in GOLD⁵⁵ and allowing full flexibility for the ligand, including flipping of amide bonds. The search sphere was centered on the side chain (CD1) of Ile 7.39, and expanded with a radius of 15 Å, thus ensuring a generous enough search space comprising the

antagonist binding site experimentally determined for adenosine receptors.⁴⁶ The criterion for the selection of docking poses was based on Chemscore ranking and the population of the solutions (according to a clustering criteria of 1 Å).

Geometrical Optimization. Each docking pose was refined by partial energy minimization of the binding site with MOE.⁷⁴ The site was selected as any atom within a distance of 4.5 Å around the ligand and OPLS-AA parameters were used in combination with GBSA model for continuum solvent representation. The convergence criterion for the steepest descent algorithm was set to 0.01 rmsd.

3D-QSAR. The conformations of the compounds obtained in the molecular docking step were used to generate a 3D-QSAR model with the software Pentacle v1.1.⁷⁵ This software allows the computation of the second generation of GRIND Independent Descriptors (GRIND-2). This family of molecular descriptors, which are widely used in QSAR studies, are generated in a three-step fashion: (i) computation of molecular interaction fields (MIF) with different Grid probes,⁷⁶ (ii) selection of the most relevant MIF nodes, and (iii) encoding of the descriptors as alignment-independent vectors of node pairs, obtaining the so-called correlograms.⁶¹ The advantages of GRIND-2 include the use of AMANDA as a new discretization algorithm for the identification of "hot spots" (most relevant MIF nodes)⁵⁰ and a new method for encoding descriptors into alignment-free vectors called CLACC.⁵¹ This method detects consistency in the computed variables, ensuring that a given vector on the correlogram corresponds to the description of same pharmacophoric property within the series. In this work, the MIF were computed using default values (i.e., GRID probes: DRY, O, N1, TIP; 0.5 Å grid step; dynamic parametrization), discretization was carried out with default AMANDA parameters and the CLACC encoding of the variables was generated on the basis of the docking alignment ("use CLACC for alignment" = false) with strict options, meaning that any variable that is not consistent in the series is removed ("Remove non-consistent couples" = true). Two rounds of fractional factorial design (FFD) were applied for the selection of the most relevant variables in the model. The model generated was "saved for predictions" through the corresponding menu option in Pentacle. Thereafter, the designed molecules were docked in the A₃ receptor as explained above, and further imported in the Pentacle software, but using the generated model as a template for the prediction of activities on the A₃ receptor.

Water Analysis. The prediction of energetically favorable regions for structural water molecules in the binding site of the hA₃AR model was carried out with the program Grid,⁷⁶ using the following parameters: OH2 probe, all program directives on their default values except for LEAU = 2 and NPLA = 2. Further refinement of the position of predicted water molecules was done with the module FilMap, as implemented in the Grid software.

Acknowledgment. This work has been developed in the frame of the Red Gallega de Investigación y Desarrollo de Medicamentos (REGID) and was financially supported by the Fondo Europeo de Desarrollo Social (FEDER) and the Galician Government (Projects 09CSA016234PR and PS09/63). E.S. is the recipient of a Consolidation Group Research Grant from the Consellería de Educación (Xunta de Galicia). E.S., H.G.d.T., and A.C. are researchers of the Isidro Parga

Pondal program (Xunta de Galicia, Spain). J.B. is researcher of the Isabel Barreto program (Xunta de Galicia, Spain). D.R. is recipient of a PFIS grant from the Instituto de Salud Carlos III (Ministerio de Ciencia e Innovación, Spain). We are grateful to Dr. Manuel Pastor for assistance in the use of the Pentacle software and helpful discussions.

Supporting Information Available: Experimental details for the synthesis of compounds described, the spectroscopic, spectrometric, and elemental analysis data of all compounds prepared as well as additional molecular modeling details. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- (1) Fredholm, B. B.; Arslan, G.; Halldner, L.; Kull, B.; Shulte, G.; Wasserman, W. Structure and function of adenosine receptors and their genes. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2000**, *362*, 364–374.
- (2) Murphree, L. J.; Linden, J. Adenosine receptors. *Encyclopaedia Biol. Chem.* **2004**, *1*, 34–39.
- (3) Jacobson, K. A.; Knutsen, L. J. S. P1 and P2 purine and pyrimidine receptor ligands. In *Purinergic and Pyrimidine Signalling (Handbook of Experimental Pharmacology)*; Abbraccio, M., Williams, M., Eds.; Springer: Berlin, 2001; Vol. 151/1, pp 129–175.
- (4) Fredholm, B. B.; Ijzerman, A. P.; Jacobson, K. A.; Koltz, K. N.; Linden, J. International union of pharmacology XXV. Nomenclature and classification of adenosine receptors. *Pharmacol. Rev.* **2001**, *53*, 527–552.
- (5) Baraldi, P. G.; Aghadeh, M.; Tabrizi, M. A.; Gessi, S.; Borea, P. A. Adenosine receptor antagonists: translating medicinal chemistry and pharmacology into clinical utility. *Chem. Rev.* **2008**, *108*, 238–263.
- (6) Jacobson, K. A.; Gao, Z. Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discovery* **2006**, *5*, 247–264.
- (7) Moro, S.; Gao, Z. G.; Jacobson, K. A.; Spalluto, G. Progress in the pursuit of therapeutic adenosine receptor antagonists. *Med. Res. Rev.* **2006**, *26*, 131–159.
- (8) Zhou, G. L.; Olah, M. E.; Johnson, R. A.; Stiles, G. L.; Clivelli, O. Molecular cloning and characterization of an adenosine receptor: the A₃ adenosine receptor. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 7432–7436.
- (9) Van Schaick, E. A.; Jacobson, K. A.; Kim, H. O.; Ijzerman, A. P.; Danhof, M. Hemodynamic effects and histamine release elicited by the selective adenosine A₃ receptor agonist CI-IB-MECA in conscious rats. *Eur. J. Pharmacol.* **1996**, *308*, 311–314.
- (10) Hannon, J. P.; Pfannkuche, H. J.; Fozard, J. R. A role for mast cells in adenosine A₃ receptor-mediated hypotension in the rat. *Br. J. Pharmacol.* **1995**, *115*, 945–952.
- (11) Fishman, P.; Bar-Yehuda, S. Pharmacology and therapeutic applications of A₃ receptor subtypes. *Curr. Top. Med. Chem.* **2003**, *3*, 463–469.
- (12) Muller, C. E. A₃ Adenosine receptor antagonists. *Mini Rev. Med. Chem.* **2001**, *1*, 417–427.
- (13) Baraldi, P. G.; Tabrizi, M. A.; Fruttarolo, F.; Borevo, A.; Avitabile, B.; Preti, D.; Romagnoli, R.; Merighi, S.; Gessi, S.; Varani, K.; Borea, P. A. Recent developments in the field of A₃ adenosine receptor antagonists. *Drug Dev. Res.* **2003**, *58*, 315–329.
- (14) Baraldi, P. G.; Fruttarolo, F.; Tabrizi, M. A.; Romagnoli, R.; Preti, D.; Carrion, M.; Iaconinoto, A.; Borea, P. A. Recent improvements in the field of A₃ adenosine receptor ligands. *Exp. Opin. Ther. Pat.* **2005**, *15*, 1507–1519.
- (15) Muller, C. E. Medicinal chemistry of adenosine A₃ receptor ligands. *Curr. Top. Med. Chem.* **2003**, *3*, 445–462.
- (16) Moro, S.; Deflorian, F. Novel strategies for the design of new potent A₃ antagonists: an update. *Curr. Med. Chem.* **2006**, *13*, 639–645.
- (17) Baraldi, P. G. A₃ ligands, history and perspectives. *Med. Res. Rev.* **2000**, *20*, 103–128.
- (18) Van Muijilwijk-Kozen, J. E.; Timmerman, H.; Ijzerman, A. P. The adenosine A₃ receptor and its ligands. *Prog. Med. Chem.* **2001**, *38*, 61–113.
- (19) Stone, T. W. Purines and neuroprotection. *Adv. Exp. Med. Biol.* **2002**, *513*, 249–280.
- (20) Beaven, M. A.; Ramkumar, V.; Hydar, A. Adenosine A₃ receptors in mast cells. *Trends Pharmacol. Sci.* **1994**, *15*, 13–14.

- (12) Ramkumar, V.; Stiles, G. L.; Beaven, M. A.; Ali, H. The A_3 adenosine receptor is the unique adenosine receptor which facilitates release of allergenic mediators in mast cells. *J. Biol. Chem.* **1993**, *268*, 16887–16890.
- (22) Akkari, R.; Burbil, J. C.; Hockemeyer, J.; Mueller, C. E. Recent progress in the development of adenosine receptor ligands as antiinflammatory drugs. *Curr. Top. Med. Chem.* **2006**, *6*, 1375–1399.
- (23) Von Lubitz, D. K. J. E.; Carter, M. F.; Deutsch, S. I.; Lin, R. C. S.; Mastropaola, J.; Meshulam, Y.; Jacobson, K. A. The effects of adenosine receptor stimulation on seizures in mice. *Eur. J. Pharmacol.* **1995**, *275*, 23–29.
- (24) Nieber, K.; Lewerenz, A.; Hentschel, S.; Vissienon, Z. Adenosine A_1 and A_3 receptors: Neuroprotective targets with perspectives. *Bioforum* **2002**, *25*, 237–240.
- (25) Forsythe, P.; Ennis, M. Adenosine, mast cells and asthma. *Inflammation Res.* **1999**, *48*, 301–307.
- (26) Brown, R. A.; Spina, D.; Page, C. P. Adenosine receptors and asthma. *Br. J. Pharmacol.* **2008**, *153*, S446–S456.
- (27) Brambilla, R.; Cattabeni, F.; Ceruti, S.; Barbieri, D. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **2000**, *361*, 225–234.
- (28) Jacobson, K. A.; Moro, S.; Kim, Y. C.; Li, A. A_3 Adenosine receptors: protective vs damaging effects identified using novel agonists and antagonists. *Drug Dev. Res.* **1998**, *45*, 113–124.
- (29) Abbraccio, M. P.; Ceruti, S.; Brambilla, R.; Barbieri, D.; Camurri, A.; Franceschi, C.; Giammaroli, A. M.; Jacobson, K. A.; Cattabeni, F.; Malorni, W. Adenosine A_3 receptors and viability of astrocytes. *Drug Dev. Res.* **1998**, *45*, 379–386.
- (30) Baraldi, P. G.; Tabrizi, M. A.; Romagnoli, R.; Fruttarolo, F.; Merighi, S.; Varani, K.; Gessi, S.; Borea, P. A. Pyrazolo[4,3-*e*]1,2,4-triazolo[1,5-*c*]pyrimidine ligands, new tools to characterize A_3 adenosine receptors in human tumor cell lines. *Curr. Med. Chem.* **2005**, *12*, 1319–1329.
- (31) Merighi, S.; Mirandola, P.; Varani, K.; Gessi, S.; Capitani, S.; Leung, E.; Baraldi, P. G.; Tabrizi, M. A.; Borea, P. A. Pyrazolo-triazolopyrimidine derivatives sensitize melanoma cells to the chemotherapeutic drugs: taxol and vindesine. *Biochem. Pharmacol.* **2003**, *66*, 739–748.
- (32) Cosimelli, B.; Greco, G.; Ehlaro, M.; Novellino, E.; Da Settimo, F.; Taliani, S.; La Motta, C.; Bellandi, M.; Tuccinardi, T.; Martinelli, A.; Ciampi, O. Derivatives of 4-amino-6-hydroxy-2-mercaptopyrimidine as novel, potent and selective A_3 adenosine antagonists. *J. Med. Chem.* **2008**, *51*, 1764–1770.
- (33) Van Veldhoven, J. P. D.; Chang, L. C. W.; von Frijtag, J. K.; Mulder-Krieger, T.; Struense-Link, R.; Beukers, M. W.; Brussee, J.; Ijzerman, A. P. A new generation of adenosine receptor antagonists: From di- to trisubstituted aminopyrimidines. *Bioorg. Med. Chem.* **2008**, *16*, 2741–2752.
- (34) Gillespie, R. J.; Bamford, S. J.; Clay, A.; Gaur, S.; Haymes, T.; Jackson, P. S.; Jordan, A. M.; Klenke, B.; Leonardi, S.; Liu, J.; Mansell, H. L.; Ng, S.; Saadi, M.; Simmonite, H.; Stratton, G. C.; Todd, R. S.; Williamson, D. S.; Yule, I. A. Antagonists of the human A_2A receptor. Part 6: Further optimization of pyrimidine-4-carboxamides. *Bioorg. Med. Chem.* **2009**, *17*, 6590–6605.
- (35) Zhang, X.; Tellew, J. E.; Luo, Z.; Moorjani, M.; Lin, E.; Lanier, M. C.; Chen, Y.; Williams, J. P.; Saunders, J.; Lechner, S. M.; Markison, S.; Joswig, T.; Petroski, R.; Piercey, J.; Kargo, W.; Malany, S.; Santos, M.; Gross, R. S.; Wen, J.; Jalali, K.; O'Brien, Z.; Stotz, C. E.; Crespo, M. I.; Diaz, J. L.; Slee, D. H. Lead optimization of 4-acetylamino-2-(3,5-dimethylpyrazol-1-yl)-6-pyridylpyrimidines as A_{2A} adenosine receptor antagonists for the treatment of parkinson's disease. *J. Med. Chem.* **2008**, *51*, 7099–7110.
- (36) Vidal, B.; Nueda, A.; Esteve, C.; Domenech, T.; Benito, S.; Reinoso, R. F.; Pont, M.; Calbet, M.; López, R.; Cadavid, M. I.; Loza, M. I.; Cárdenas, A.; Godesart, N.; Beleta, J.; Warreilow, G.; Ryder, H. Discovery and characterization of 4'-(2-furyl)-N-pyridin-3-yl-4,5'-bipyrimidin-2'-amine (LAS38096), a potent, selective, and efficacious A_{2B} adenosine receptor antagonist. *J. Med. Chem.* **2007**, *50*, 2732–2736.
- (37) Kuefner-Muehl, U.; Scheuplein, S. W.; Pohl, G.; Gaida, W.; Lehr, E.; Mierau, J.; Meade, C. J. M. Triazines with an adenosine antagonistic effect. PCT Int. Appl. WO 1999011633.
- (38) Morizzo, E.; Capelli, F.; Lenzi, O.; Catarzi, D.; Varano, C.; Filacchioni, G.; Vicenzi, F.; Varani, K.; Borea, P. A.; Colotta, V.; Moro, S. Scouting human A_3 adenosine receptor antagonist binding mode using a molecular simplification approach: from triazoloquinoxaline to a pyrimidine skeleton as a key study. *J. Med. Chem.* **2007**, *50*, 6596–6606.
- (39) Chang, L. C. W.; Spanjersberg, R. F.; von Frijtag, J. K.; Mulder-Krieger, T.; van den Hout, G.; Beukers, M. W.; Brussee, J.; Ijzerman, A. P. 2,4,6-Trisubstituted pyrimidines as a new class of selective adenosine A_1 receptor antagonists. *J. Med. Chem.* **2004**, *47*, 6529–6540.
- (40) Cheong, S. L.; Dolzhenko, A.; Kachler, S.; Paoletta, S.; Federico, S.; Cacciari, B.; Dolzhenko, A.; Koltz, K. N.; Moro, S.; Spalluto, G.; Pastorin, G. The significance of 2-furyl ring substitution with a 2-(*para*-substituted) aryl group in a new series of pyrazolo-triazolo-pyrimidines as potent and highly selective A_3 adenosine receptors antagonists: new insights into structure–affinity relationship and receptor–antagonists recognition. *J. Med. Chem.* **2010**, *53*, 3361–3375.
- (41) Priego, E. M.; Pérez-Pérez, M. J.; von Frijtag, J. K.; de Vries, H.; Ijzerman, A. P.; Camarasa, M. J.; Martín-Santamaría, S. Selective human adenosine A_3 antagonists based on pyrido[2,1-*f*]purine-2,4-diones: novel features of A_3 antagonist binding. *ChemMedChem* **2008**, *3*, 111–119.
- (42) Baraldi, P. G.; Tabrizi, M. A.; Preti, D.; Bovero, A.; Fruttarolo, F.; Romagnoli, R.; Zaid, N. Z.; Moorman, A. R.; Varani, K.; Borea, P. A. New 2-arylpyrazolo[4,3-*c*]quinoline derivatives as potent and selective human A_3 adenosine receptor antagonists. *J. Med. Chem.* **2005**, *48*, 5001–5008.
- (43) Van Muijlwijk-Kozen, J. E.; Timmerman, H.; Vollinga, R. C.; von Drabbe, J. F.; de Groote, M.; Visser, S.; Ijzerman, A. P. Thiazole and thiazidazole analogues as a novel class of adenosine receptor antagonists. *J. Med. Chem.* **2001**, *44*, 749–762.
- (44) Michino, M.; Abola, E.; Brooks, C. L.; Dixon, J. S.; Moul, J.; Stevens, R. C. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **2009**, *8*, 455–463.
- (45) Bondavalli, F.; Botta, M.; Bruno, O.; Ciacci, A.; Corelli, F.; Fossa, P.; Lucacchini, A.; Manetti, F.; Martini, C.; Menozzi, G.; Mosti, L.; Ranise, A.; Schenone, S.; Tafi, A.; Trincavelli, M. L. Synthesis, molecular modeling studies, and pharmacological activity of selective A_1 receptor antagonists. *J. Med. Chem.* **2002**, *45*, 4875–4887.
- (46) Moro, S.; DeLorain, F.; Spalluto, G.; Pastorin, G.; Cacciari, B.; Kim, S.; Jacobson, K. A. Demystifying the three dimensional structure of G protein-coupled receptors (GPCRs) with the aid of molecular modelling. *Chem. Commun.* **2003**, *24*, 2949–2956.
- (47) Jaakola, V.; Griffith, M. T.; Hanson, M. A.; Cherezov, V.; Chien, E. Y. T.; Lane, J. R.; Ijzerman, A. P.; y Stevens, R. C. The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **2008**, *322*, 1211–1217.
- (48) Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. Structure-based discovery of beta 2-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 6843–6848.
- (49) Lenzi, O.; Colotta, V.; Catarzi, D.; Varano, F.; Poli, D.; Filacchioni, G.; Varani, K.; Vicenzi, F.; Borea, P. A.; Paoletta, S.; Morizzo, E.; Moro, S. 2-Phenylpyrazolo[4,3-*d*]pyrimidin-7-one as a new scaffold to obtain potent and selective human A_3 adenosine receptor antagonists: new insights into the receptor-antagonist recognition. *J. Med. Chem.* **2009**, *52*, 7640–7652.
- (50) Durán, A.; Martínez, G. C.; Pastor, M. Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in Molecular Interaction Fields. *J. Chem. Inf. Model.* **2008**, *48*, 1813–1823.
- (51) Duran, A. Ph.D. Thesis, University Pompeu Fabra, Barcelona, 2010. <http://www.tesisenxarxa.net/TDX-0422110-094351>
- (52) Yaziji, V.; Coelho, A.; El Maatougui, A.; Brea, J.; Loza, M. I.; Garcia-Mera, X.; Sotelo, E. Divergent solution-phase diarylpyrimidine libraries as selective A_3 adenosine receptor antagonists. *J. Comb. Chem.* **2009**, *11*, 519–522 and references cited there.
- (53) Bosch, M. P.; Campos, F.; Niubo, I.; Rosell, G.; Diaz, J. L.; Brea, J.; Loza, M. I.; Guerrero, A. Synthesis and biological activity of new potential agonists for the human adenosine A_{2A} receptor. *J. Med. Chem.* **2004**, *47*, 4041–4053.
- (54) Leff, P.; Dougall, I. G. Further concerns over Cheng-Prusoff analysis. *Trends Pharmacol. Sci.* **1993**, *14*, 110–112.
- (55) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609–623.
- (56) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
- (57) Jaakola, V. P.; Lane, J. R.; Lin, J. Y.; Katritch, V.; Ijzerman, A. P.; Stevens, R. C. Ligand binding and subtype selectivity of the human A_{2A} adenosine receptor: identification and characterization of essential amino acid residues. *J. Biol. Chem.* **2010**, *285*, 13032–13044.
- (58) Gao, Z. G.; Chen, A.; Barak, D.; Kim, S. K.; Muller, C. E.; Jacobson, K. A. Identification by site-directed mutagenesis of residues involved in ligand recognition and activation of the

- human A₃ adenosine receptor. *J. Biol. Chem.* **2002**, *277*, 19056–19063.
- (59) Kim, J.; Wess, J.; van Rhee, A. M.; Schöneberg, T.; Jacobson, K. A. Mutagenesis identifies residues involved in ligand recognition in the human A_{2A} adenosine receptor. *J. Biol. Chem.* **1995**, *270*, 13987–13997.
- (60) Kim, S. K.; Gao, Z. G.; Rompaey, P. V.; Gross, A. S.; Chen, A.; Calenbergh, S. V.; Jacobson, K. A. Comparison of the binding domains of A_{2A} agonists and antagonists. *J. Med. Chem.* **2003**, *46*, 4847–4859.
- (61) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRIND-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.
- (62) Tucker, A. L.; Robeva, A. S.; Taylor, H. E.; Holeton, D.; Bockner, M.; Lynch, K. R.; Linden, J. A₁ adenosine receptors. Two amino acids are responsible for species differences in ligand recognition. *J. Biol. Chem.* **1994**, *269*, 27900–27906.
- (63) Ivanov, A. A.; Barak, D.; Jacobson, K. A. Evaluation of homology modeling of G-protein-coupled receptors in light of the A(2A) adenosine receptor crystallographic structure. *J. Med. Chem.* **2009**, *52*, 3284–3292.
- (64) Benderitter, P.; Xavier de Araujo, J.; Schmitta, M.; Bourguignon, J. J. 2-Amino-6-iodo-4-tosyloxypyrimidine: a versatile key intermediate for regioselective functionalization of 2-aminopyrimidines in 4- and 6-positions. *Tetrahedron* **2007**, *63*, 12465–12470.
- (65) Gong, B.; Hong, F.; Kohm, C.; Jenkins, S.; Tulinsky, J.; Bhatt, R.; Vries, P.; Singer, J. W.; Klein, P. Synthesis, SAR, and antitumor properties of diamino-C₅Ndiarylpurimidine positional isomers: inhibitors of lysophosphatidic acid acyltransferase- β . *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2303–2308.
- (66) Swarbrick, M. E.; Beswick, P. J.; Gleave, R. J.; Green, R. H.; Bingham, S.; Bountra, C.; Carter, M. C.; Chambers, L. J.; Chessell, I. P.; Clayton, N. M.; Collins, S. D.; Corfield, J. A.; Hartley, C. D.; Kleanthous, S.; Lambeth, P. F.; Lucas, F. S.; Mathews, N.; Naylor, A.; Page, L. W.; Payne, J. J.; Pegg, N. A.; Price, H. S.; Skidmore, J.; Stevens, A. J.; Stocker, R.; Stratton, S. C.; Stuart, A. J.; Wiseman, J. O. Identification of [4-[4-(methylsulfonyl)phenyl]-6-(trifluoromethyl)-2-pyrimidinyl] amines and ethers as potent and selective cyclooxygenase-2 inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4504–4508.
- (67) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Res.* **1994**, *22*, 4673–4680.
- (68) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (69) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- (70) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, *35*, W375–383.
- (71) Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.
- (72) *Macromodel*, version 9.7; Schrödinger, LLC: New York, NY, 2009.
- (73) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
- (74) *MOE (Molecular Operative Environment)*, version 2009.10; Chemical Computing Group Inc. Montreal, Canada; 2009.
- (75) *Pentacle*, version 1.1; Molecular Discovery, Ltd.: Perugia, Italy, 2009.
- (76) Goodford, P. J. A computational procedure for determining energetically favorable binding sites of biological important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

Paper IV

Rodríguez D., Piñeiro Á. and Gutiérrez-de-Terán H.

Molecular Dynamics Simulations Reveal Insights into Key Structural
Elements of Adenosine Receptors

Biochemistry, **2011** 50(19):4194-208

Molecular Dynamics Simulations Reveal Insights into Key Structural Elements of Adenosine Receptors

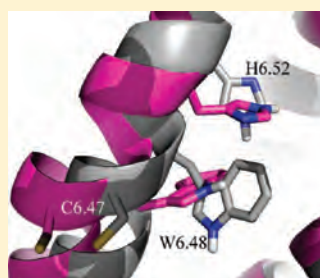
David Rodríguez,[†] Ángel Piñeiro,[‡] and Hugo Gutiérrez-de-Terán^{*†}

[†]Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago (CHUS), planta-2, A Choupana, s/n E-15706 Santiago de Compostela, Spain

[‡]Soft Matter and Molecular Biophysics Group, Department of Applied Physics, University of Santiago de Compostela, Campus Vida s/n, E-15782 Santiago de Compostela, Spain

S Supporting Information

ABSTRACT: The crystal structure of the human A_{2A} adenosine receptor, a member of the G protein-coupled receptor (GPCR) family, is used as a starting point for the structural characterization of the conformational equilibrium around the inactive conformation of the human A₂ (A_{2A} and A_{2B}) adenosine receptors (ARs). A homology model of the closely related A_{2B}AR is reported, and the two receptors were simulated in their apo form through all-atom molecular dynamics (MD) simulations. Different conditions were additionally explored in the A_{2A}AR, including the protonation state of crucial histidines or the presence of the cocrystallized ligand. Our simulations reveal the role of several conserved residues in the ARs in the conformational equilibrium of the receptors. The “ionic lock” absent in the crystal structure of the inactive A_{2A}AR is rapidly formed in the two simulated receptors, and a complex network of interacting residues is presented that further stabilizes this structural element. Notably, the observed rotameric transition of Trp6.48 (“toggle switch”), which is thought to initiate the activation process in GPCRs, is accompanied by a concerted rotation of the conserved residue of the A₂ARs, His6.52. This new conformation is further stabilized in the two receptors under study by a novel interaction network involving residues in transmembrane (TM) helices TMs (Asn5.42) and TM3 (Gln3.37), which resemble the conformational changes recently observed in the agonist-bound structure of β-adrenoreceptors. Finally, the interaction between Glu1.39 and His7.43, a pair of conserved residues in the family of ARs, is found to be weaker than previously thought, and the role of this interaction in the structure and dynamics of the receptor is thoroughly examined. All these findings suggest that, despite the commonalities with other GPCRs, the conformational equilibrium of ARs is also modulated by specific residues of the family.



G protein-coupled receptors (GPCRs), the largest family of membrane receptors in the human genome, are transmembrane proteins in charge of the transduction of signals across cellular membranes.¹ A delicate conformational equilibrium between extreme active and inactive forms of these receptors is behind the signal transduction, which is modulated toward activation by the binding of their respective natural agonist.² From a structural point of view, GPCRs share a conserved topology: a helical bundle consisting of seven transmembrane (TM) helices spanning the membrane, which are connected by three extracellular (EL) and three intracellular (IL) loops. Recent advances in structural biology have revealed that the largest differences among GPCRs are located in their extracellular halves, where ligand binding occurs.³ A general phylogenetic classification of the GPCR superfamily⁴ shows well-defined clusters, which can be related to the chemical nature of the receptor's natural ligand.

Adenosine receptors (ARs) make up one family of GPCRs, which mediate the important extracellular role of the nucleoside adenosine. Four AR subtypes exist, namely, A₁, A_{2A}, A_{2B}, and A₃,

with different signaling functions and tissue distributions. The pharmacological interest in ARs as drug targets is widely recognized because of the variety and importance of the physiological functions mediated by them, as well as the associated pathologies.⁵ Pharmacological applications of AR chemical modulators include several inflammatory processes, where the role of the A_{2A}AR and A₃AR is generally accepted,⁶ or several respiratory pathological events such as allergic asthma, where the nonselective AR antagonist theophylline is one of the main treatments, although more selective A_{2B}AR antagonists are in clinical trials.⁷ The use of adenosine itself in the treatment of certain arrhythmias and the anti-infarct effect of other adenosine agonists is mediated by the A₁AR,⁸ while the A_{2A}AR is involved in vascular diseases such as hypertension or atherosclerosis.⁹ ARs are also involved in neural regulation in the central nervous system, as exemplified by the well-known AR antagonist caffeine.¹⁰

Received: January 21, 2011

Revised: April 11, 2011

Published: April 11, 2011

The recent discovery of A_{2A}AR–dopamine D₂ receptor heterodimers in ganglionic neurons led to increased interest in the A_{2A}AR as a target for Alzheimer's disease.¹¹

Given the high degree of biomedical relevance of the AR family members, it is not difficult to understand the breakthrough that meant the release of the hA_{2A}AR crystal structure at a resolution of 2.6 Å.¹² The structure represents one of the inactive states of the receptor with the subtype selective high-affinity antagonist ZM241385 bound to it. As in the case of the human β₂-adrenergic receptor (hβ₂),¹³ stabilization of the receptor was achieved by substitution of the third intracellular loop (IL3) with the T4 bacteriophage lysozyme (T4L fusion protein strategy). The long C-terminus (Ala³¹⁷–Ser⁴¹²) was deleted, while the first two residues of the N-terminus and seven residues (149–155) of the second extracellular loop (EL2) were not determined,¹⁴ which probably indicated the high mobility of these regions. Despite these limitations, the structure of the A_{2A}AR–ZM241385 complex has revealed important particular features of the AR family and offers a new scenario for the study of the ARs, considering the high degree of sequence identity within their members. As recently reviewed, the helical bundle orientation and packing is similar among the GPCRs with known structures.³ The major differences are located in the loop regions and within the binding site cavity. It is precisely the A_{2A}AR that shows the most divergence: the conformations of EL2 and EL3 are constrained because of the presence of three extra disulfide bonds, besides the one conserved in all GPCRs between EL2 and TM3. Additionally, the binding site not only is more shifted toward helices TM6 and TM7, as previously suggested by site-directed mutagenesis and molecular modeling studies of ARs,¹⁵ but also includes important direct interactions of residues at EL2 with the cocrystallized ligand.

Until very recently, GPCR crystal structures had been determined with either an antagonist or an inverse agonist bound, thus stabilizing an inactive conformation of the receptor. An exception to this rule was the opsin, the ligand-free active form of rhodopsin,¹⁶ followed by very recent structures of β-adrenergic receptors in different activelike conformations.^{17–19} The elucidation of the structure and dynamic pathways, connecting the ensemble of conformational states of a GPCR,² is a question of major interest. According to the currently accepted GPCR activation theories, two structural elements play a key role in the activation process: (i) a salt bridge formed by residue Arg3.50 (see Materials and Methods for residue numbering), belonging to the (E/D)RY motif, and Glu6.30 known as an “ionic lock”^{20,21} and (ii) the “toggle switch”, a term that refers to the rotameric transition of Trp6.48 in TM6.²² While some authors propose that a connection should exist between these two events,²³ biophysical experiments with the β₂-adrenergic receptor suggest that the disruption of the ionic lock and the activation of the rotameric toggle switch are not tightly coupled.²⁴ The ionic lock is present in the inactive crystal structure of bovine rhodopsin but not in the experimental structure representing a model of the active form (opsin),¹⁶ which has led to the idea that the rupture of this salt bridge is a switch in the activation process. Whereas the very recently determined structure of the inactive D₃ dopamine receptor (D₃DR) shows a formed ionic lock,²⁵ a disrupted TM3–TM6 ionic lock is observed in the rest of the GPCRs crystallized in the inactive form (hA_{2A}AR and β₁- and hβ₂-adrenergic receptors).¹ A comparison of active and inactive forms of (rhod)opsins reveals that the degree of TM6 bending is lower in the active conformation. This is accompanied by an

extension of the TMS and TM6 helical segments in their intracellular part. Such observations are in agreement with the activation theories proposed for the family of adrenergic receptors²² and further confirmed by the very recently determined crystal structure of the nanobody-stabilized active state of hβ₂.¹⁹ Interestingly, the partial and full agonist-bound structures of the β-adrenoreceptors provide new insights into the dynamic role of the connections among TM3, TMS, and TM6 in the initial steps of the activation process.^{17,18}

With the availability of crystal structures, there has been a renaissance of GPCR computational and structural studies. Computational techniques like molecular dynamics (MD) simulations,^{26–30} protein–protein docking,³¹ and virtual screening³² have gone beyond the static crystallographic structures of the β-adrenergic receptors and have allowed for a more comprehensive exploration of the functionality of this family of GPCRs. In particular, MD simulations have provided important insights into delicate conformational changes such as the ionic lock and its microenvironment in the β-adrenergic receptors.^{26–29} Additionally, computational studies of rhodopsin employing the LITICON method,³³ prior to the release of the activated opsin structure, and recent metadynamics simulations of the same system³⁴ already predicted the conformational events observed in the rhodopsin activation process.

Two years after the release of the A_{2A}AR structure, some effort has been spent on this, including the identification of novel chemotypes using structure-based virtual screening,³⁵ the postulation of the stabilization effect of cholesterol molecules in the hA_{2A}AR,³⁶ and the influence of the saline concentration in the formation of the ionic lock,³⁰ both items studied by MD simulations on this particular receptor. Despite the importance of these initial efforts, a deeper structural and dynamic exploration of the ARs is clearly needed to gain comprehensive insight into this particular family of GPCRs. In this work, all-atom molecular dynamics simulations of the ligand-free structures of the two human A₂ receptors (hA_{2A}AR and hA_{2B}AR) are performed in a hydrated membrane model. Moreover, simulations of the A_{2A}AR evaluating different variables for further exploring the results of the apo form of the receptor are conducted, including the consideration of the cocrystallized antagonist ZM241385. With this approach, conserved motifs in the GPCR superfamily, identified by sequence^{37,38} and recent comparative structural analysis,^{3,39} are deeply analyzed, and their role in the conformational equilibrium of ARs is recognized. In addition to the formation of the ionic lock and the characterization of the connecting residues among transmembrane helices TM1, TM2, and TM7, we report a novel network of residues connecting TM3, TMS, and TM6 that we propose to be an important step for the activation process. The structural role of several particular residues of the AR family arises, which is discussed in the context of available experimental data.

■ MATERIALS AND METHODS

Residue Numbering. The general GPCR numbering scheme proposed by Ballesteros and Weinstein⁴⁰ was adopted throughout this work because it allows a proper comparison between different receptors. Briefly, every residue is labeled as X.YY, where number X corresponds to the TM helix (from 1 to 7) and YY is a correlative number starting from the most conserved residue in each helix (to which a YY value of 50 is assigned). Additionally, the absolute receptor sequence number is indicated with a

superscript when considered necessary (i.e., loop regions), following the corresponding Swiss-Prot sequences (P29274 for A_{2A}AR and P29275 for A_{2B}AR).

Molecular Modeling of Adenosine Receptors. The structure of the A_{2A}AR in a complex with its potent inhibitor ZM241385 was retrieved from the Protein Data Bank (PDB) (entry 3EML). The missing regions of the receptor were modeled by using Modeller version 9.4.⁴¹ These include a fragment of EL2 (¹⁴⁹PKEGKNH¹⁵⁵), undetermined in the crystal structure, and the IL3 segment (²⁰⁹KQMESQLPGERA²²¹), which was substituted with the T4 bacteriophage lysozyme in that structure. The other two missing regions were not considered in this work, i.e., a small N-terminal fragment (¹MP²) and the 95-residue C-terminus. Fifteen initial models of the missing regions were built. The top five models, according to DOPEHR scoring, were selected for further refinement with the LoopModel routine.⁴² Thus, a pool of 10 models per input structure were generated. One model per input structure was selected again according to DOPEHR scoring, and a final selection among the five candidates was done by a combination of visual inspection and quality analysis using Procheck⁴³ and Molprobit.⁴⁴ The side chain conformations of all Asn, Gln, and His residues were assessed with Molprobit⁴⁴ and PDB2PQR⁴⁵ prediction servers. The Protein Preparation Wizard utility (Schrödinger LLC, New York, NY) was employed to add hydrogens and initially define protonation states of titratable residues and Ser/Thr rotamer assignments. A deeper exploration of the protonation states of Glu1.39–His7.43 and Glu¹⁶⁹(5.30)–His²⁶⁴(7.29) residue pairs was performed with MCCE.⁴⁶ Briefly, this software generates side chain conformers for the residues under study, performs continuum electrostatics calculations with DelPhi⁴⁷ as the Poisson–Boltzmann equation solver, and performs a final titration simulation employing Monte Carlo sampling. Twelve different rotamers were generated for the aforementioned residues and those within 0.4 nm of them. A slab of neutral atoms surrounding the protein was added to mimic the effect of a cellular membrane.⁴⁸ Default parameters were employed for the rest of the directives.

A partial energy minimization of the modeled loops was performed on the selected structure with MacroModel,⁴⁹ using the Polak–Ribiere Conjugate Gradient (PRCG) with a convergence criterion of 0.05, and the OPLS2005 force field^{50,51} in combination with the GBSA solvation model. The modeled regions were unrestrained, applying a positional harmonic constraint (2×10^4 kJ mol⁻¹ nm⁻²) to the rest of EL2 and any other residue within 2 Å of modeled regions, while the rest of the protein was completely frozen.

The hA_{2B}AR structure was obtained by homology modeling using Modeller with the hA_{2A}AR as a template. First, a multiple-sequence alignment of multispecies A_{2A}AR and A_{2B}AR receptors was generated with ClustalX2.0,⁵² using the PAM250 substitution matrix, and open and elongation gap penalties of 10 and 0.05, respectively. In analogy with the A_{2A}AR structure, the first three residues of the N-terminus and the last 13 residues of the C-terminus were discarded for the modeling. The modeling protocol was adapted from that employed by some of us in the GPCRdock2008 competition:⁵³ 15 initial models were obtained using standard Modeller parameters, selecting the top five models on the basis of the DOPEHR scoring function to be further refined by means of the Molprobit server.⁴⁴ The best model was selected using a compromise between the DOPEHR scoring and the geometrical quality assessment of Procheck and used as a

starting point for a loop optimization procedure with the Loop-Model routine of Modeller. Only the loop regions with no secondary structure were subjected to this refinement step (i.e., the three extracellular loops, excluding the short helical fragment ¹⁷³FENV¹⁷⁶ of EL2, and IL3) and 15 models were generated. The final selection was performed using criteria identical to those used in the previous step: Procheck stereochemical quality and the DOPEHR energetic ranking. The addition of hydrogens, refinement, and optimization of the geometry of the loop regions using energy minimization procedures were performed as described above for the hA_{2A}AR.

Insertion into the Membrane and Molecular Dynamics Simulations. An automated protocol for the insertion of the GPCR structures into an explicit membrane model, using a combination of ad hoc Linux shell scripts and GROMACS utilities, was developed. Briefly, the protocol consists of the following steps. (i) The GPCR structure (hA_{2A} or hA_{2B}ARs in our case) is introduced into the center of a hexagonal prism-shaped box with the receptor TM helices and the box symmetry axes parallel to the z direction, with every side box walls at a minimum of 2.0 nm and every top and bottom box walls at a minimum of 0.8 nm of any atom. (ii) A hydrated POPC gel-phase bilayer with exactly the same dimensions as the box of the previous step, with the hydrophobic phase parallel to the x–y plane and located at half of the z dimension, is created from a larger bilayer pre-equilibrated at 260 K, just by removing the excess of lipid and water molecules. (iii) The resulting bilayer was copied on the receptor box. (iv) The water and lipid molecules overlapping any protein atom were removed. (v) Cl⁻ ions are introduced to neutralize the total charge of the receptor at physiological pH (from 6 to 12, depending on the receptor and protonation state considered for the histidine residues). The abovementioned gel phase instead of a liquid-crystalline bilayer was employed to optimize the packing of lipids around the AR model. The final systems consist of approximately 14000 water and 200 lipid molecules, in addition to the Cl⁻ ions and the receptor structure.

MD simulations were performed by applying the recently published half-ε double-pairlist method⁵⁴ to make compatible the so-called Berger united atom parameters^{55,56} used for the lipids with the OPLS-AA force field^{50,51} employed for the receptor and ions. Simple point charge (SPC) water molecules⁵⁷ were included to solvate the systems. All the simulations were performed using periodic boundary conditions in the three spatial dimensions with hexagonal prism-shaped boxes. This geometry allows us to optimize the use of the solvent with respect to the rectangular prism, by saving 13.4% of the volume for the same distance to the periodic images, and to increase the isotropy of the unavoidable interaction of the receptor with its periodic images. A semiisotropic Parrinello–Rahman barostat^{58,59} was employed to maintain the pressure independently in the x–y plane and in the z direction at 1 bar with a coupling constant of 2 ps. The isothermal compressibility constant was 4.5×10^{-5} bar⁻¹. The temperature was kept constant at 310 K using a Nose–Hoover thermostat^{58,59} by coupling independently the lipid molecules, the protein, and the water–ion groups with a common period of 0.1 ps. A cutoff of 1.2 nm was employed for the Lennard–Jones potential. The long-range interactions were calculated using the particle mesh Ewald method^{60,61} with a 1.2 nm real space cutoff, a 0.15 nm space grid, and a fourth-order B-spline interpolation scheme to optimize the computational performance, giving a PME load of ~25%. The initial velocities of the atoms were randomly assigned to produce a Maxwell

distribution corresponding to the temperature at which the simulation will be performed. The equations of motion were integrated using the leapfrog method⁶² with a 2 fs time step. The bonds lengths and H—O—H angle in water were constrained using the SETTLE algorithm,⁶³ while the LINCS algorithm⁶⁴ was used to constrain bond lengths in the protein and lipid molecules. All the simulations were performed using GROMACS^{65,66} version 4.0.5. Simulations of POPC lipid bilayers at different temperatures using the same parameters were performed to ensure that they reproduce the experimental area per lipid and the right deuterium order parameters.

An equilibration protocol was specifically designed for this system to avoid potential unfolding or distortion of protein regions because of a less exhaustive equilibration. This protocol is summarized in Table 1 and includes the following stages: (i) harmonic positional restrictions on all the heavy atoms of the protein to equilibrate the lipids and water molecules (five steps of 1 ns each, decreasing the force constant from 1000 to 200 kJ mol⁻¹ nm⁻², by 200 kJ mol⁻¹ nm⁻² in each step), (ii) harmonic positional restrictions on the α -carbon atoms of the protein to equilibrate the protein side chains with the surrounding atoms (6 ns using a force constant of 200 kJ mol⁻¹ nm⁻²), (iii) harmonic positional restrictions on the α -carbon atoms of the receptor transmembrane regions to equilibrate the loops, avoiding helix perturbations (21 ns using a force constant of 200 kJ mol⁻¹ nm⁻²), and (iv) fully unrestrained simulations [100 ns for the main simulations (see Table 2)]. After the first few nanoseconds of the equilibration stage and during the whole production trajectory, the lipids are clearly in the disordered

Table 1. Description of the Equilibration Protocol and Production Stages Performed in the Simulations of AR Models in POPC Bilayers

stage	positional harmonic restrictions	comments	time (ns)
i	on all the receptor heavy atoms	five steps with decreasing force constants from 1000 to 200 kJ mol ⁻¹ nm ⁻² ; the aim is to relax the receptor environment	5 × 1
ii	on the receptor α -carbon atoms	to relax the receptor side chains	6
iii	on the receptor α -carbon atoms of transmembrane regions	to relax the loops	21
iv	none	production trajectory	100 ^a

^a Value for the main simulations, see Table 2 for details.

Table 2. Description of the Performed Simulations, Together with the Longitudes of Their Production Phases

code(s)	description	simulated receptor	production time (ns)
R1, R2	receptor with a residue at position 7.29 in its neutral form, replicas 1 and 2	A _{2A} AR, A _{2B} AR	100
R1', R2'	receptor with residue His7.29 in its charged form, replicas 1 and 2	A _{2A} AR	90
R1''	receptor with both residues His7.29 and His7.43 in its charged form	A _{2A} AR	90
RW1	receptor with crystallographic water molecules	A _{2A} AR	90
RL1	receptor as in R1, R2, in complex with the crystallographic ligand	A _{2A} AR	30 ^a
RL1'	receptor as in R1', R2', in complex with the crystallographic ligand	A _{2A} AR	50

^a This MD simulation was not extended until 50 ns because of the high instability of the complex under these conditions.

liquid-crystal phase. By means of the scripted protocols, independent replicas for several simulations (see Table 2) were performed under equivalent conditions. All coordinates were stored every 10 ps for further analysis.

MD trajectory analyses were conducted using GROMACS utilities (see the Supplementary Methods of the Supporting Information). PyMOL (<http://www.pymol.org>) was used to prepare all molecular images and to perform protein superpositions (*super* command, on the basis of the α -carbon traces).

Inclusion of Crystallographic Water Molecules and Ligand. We performed one simulation of the A_{2A}AR [RW1 (see Table 2 and Figure S1 of the Supporting Information)], in which selected crystallographic water molecules were explicitly included: residues 502, 504, 505, 509, 512, and 565 according to PDB entry 3EML. Addition of hydrogens and refinement of associated hydrogen bonds were performed with the *Protein Preparation Wizard*, followed by energy minimization of hydrogen atoms (rmsd convergence of 3.0 nm, using the OPLS2005 force field in Macromodel).

The necessary parameters needed for the crystallographic ligand ZM241385 in the OPLSAA force field were automatically obtained with Macromodel⁴⁹ and translated to the GROMACS syntax using an ad hoc script. Two MD simulations of the A_{2A}AR–ZM241385 complex [RL1 and RL1' (see Table 2)] were conducted.

The same equilibration protocol as in the case of the apo form of the receptor was employed for these simulations, applying the same restraints to the ligand and water molecules as in the case of the side chains of the receptor.

RESULTS

Molecular Modeling. On the basis of the high degree of sequence similarity between the two human A₂ARs (59 and 70% for the whole protein and the TM domains, respectively), the crystallographic structure of the hA_{2A}AR was employed as a template to produce a homology model of the hA_{2B}AR. As expected, the main topological features of the TM region are common to both structures, the most important divergent sequences being located in the longest loops. The topology of EL2 is influenced by the four additional residues in the A_{2B}AR (37 vs 33 residues in the A_{2A}AR) and by the different arrangement of the three disulfide bonds, which were assigned to the A_{2B}AR on the basis of a family sequence alignment (data not shown). Two disulfide bridges are conserved between both receptors: the first one connects EL1 and EL2 (residues 71–159 in the A_{2A}AR and residues 72–167 in the A_{2B}AR), whereas the second one is the highly conserved disulfide bridge in the GPCR superfamily, located between EL2 (residue 166 in the A_{2A}AR and residue 171 in the A_{2B}AR) and TM3 (position 3.25).

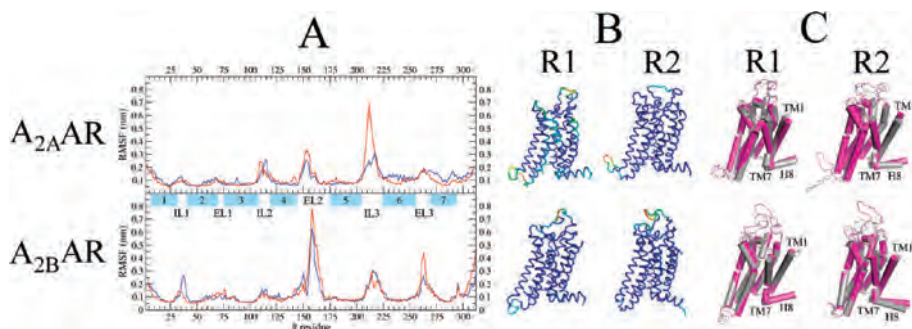


Figure 1. Dynamic properties of simulated A_2AR s. (A) Root-mean-square-fluctuation (RMSF) for each receptor under study ($A_{2A}AR$, top; $A_{2B}AR$, bottom). In each case, the blue line stands for the R1 replica and the red line accounts for replica R2. Cyan boxes denote the location of the TM helices. (B) For each MD simulation (R1, left; R2, right; $A_{2A}AR$, top; $A_{2B}AR$, bottom), the reference structure used for rmsf calculations (see the text) is represented in ribbons and colored by the gradient of the estimated β factors (from blue to red, which indicates the highest mobility). (C) Graphical representation of the PCA. The two extreme projections on the first eigenvector are represented with silver and magenta cartoons. The extracellular part of TM1 and the intracellular side of TM7 are labeled (together with C-terminal helix 8, H8). The rest of the TMs are ordered in a counterclockwise direction. The results from the different MD simulations are arranged as in panel B.

However, a third cysteine bridge is present in each receptor between nonconserved positions within the family: in the $A_{2A}AR$ this connection is established between EL2 (residue 146) and the tip of TM3 (position 3.22), offering an additional conformational constraint to the loop, while in the $A_{2B}AR$ the bridge is established between positions 154 and 166 in EL2. With respect to IL3, the modeled conformations are relatively extended in the two receptors, being four residues longer in the $A_{2A}AR$ (17 vs 13 residues in the $A_{2B}AR$). It must be noted that IL3 has been identified as an unstructured region in GPCRs;¹⁴ thus, the model quality of this region is very limited because of the absence of the cognate G protein. Finally, a fourth disulfide bridge constrains particularly the structure of EL3 in the $A_{2A}AR$. This bridge, located between positions 259 and 261, is not totally conserved in the AR family, being absent in the $A_{2B}AR$. An interaction between EL3 and EL2 is established through polar interactions of the side chains of Glu5.30 in EL2 (conserved in both $A_{2A}AR$ and $A_{2B}AR$) and position 7.29 in EL3, occupied by His²⁶⁴ in the $A_{2A}AR$ or Asn²⁶⁶ in the $A_{2B}AR$.

Geometrical and Conformational Analysis of MD Simulations. The structures obtained for the two h A_2AR s were employed as starting conformations for extensive MD simulations in an atomistic model of the membrane. Because of the crystallization conditions of the $A_{2A}AR$, bound to the ZM241385 antagonist, starting conformations of the two receptors should be considered to be in the inactive state. Most MD trajectories here discussed were obtained in the absence of any ligand, in an attempt to observe the initial stages of the activation conformational pathway, which was recently suggested in the case of the β_2 -adrenergic receptor.²⁸ This strategy is further supported by the fact that constitutive activity was described for the $A_{2A}AR$.^{67,68} Furthermore, the influence of the antagonist ligand in the dynamic behavior of the $A_{2A}AR$ was also examined (see below). To efficiently sample the conformational space and to check for reproducibility of the main structural features, two independent replicas with different random seeds were conducted for each receptor [from now on simulations R1 or R2 for either the $A_{2A}AR$ or the $A_{2B}AR$, respectively (see Table 2)].

Taking into account the fact that the entire $A_{2B}AR$ structure as well as part of the $A_{2A}AR$ was modeled, we took special care during the equilibration process, which was sufficiently long to allow the progressive relaxation of solvent molecules, loop residues, and TM regions. The 100 ns long production stage started well after a plateau for the root-mean-square deviation (rmsd) of the loops was reached (see Figure S2 of the Supporting Information). Unless the contrary is stated, the rest of the analyses will refer to the unrestrained production phase of each replica.

The backbone rmsds of the whole receptors fluctuate, as a function of time, within the values typically obtained for protein crystal structures, between 0.2 and 0.4 nm (see Figure S2 of the Supporting Information). The root-mean-square fluctuation (rmsf) values per residue (Figure 1) were determined using the most representative protein conformation as the reference structure, which was obtained by cluster analysis of each trajectory (see the Supplementary Methods of the Supporting Information). The regions with the highest mobility for the $A_{2A}AR$ were found to be IL3, EL2, and IL2. For the $A_{2B}AR$, EL2 seems to be by far the most mobile region in the two replicas, followed by EL3 and IL3. The reason for the higher flexibility of EL2 in the $A_{2B}AR$ must be found in the different arrangement of the disulfide bond in this loop: in this receptor, the disulfide bond occurs between two cysteine residues located within EL2, which consequently has more conformational freedom, whereas in the $A_{2A}AR$, this bond connects EL2 to the end of TM3, thus restraining the movement of the loop. The rmsf results fit well with the experimental data available for the $A_{2A}AR$ crystal structure, because it is precisely the unsolved tip of EL2 (residues 149–155) showing the largest fluctuations (Figure 1A,B). Also in the extracellular region, EL3 shows a higher mobility in the $A_{2B}AR$, likely because of the lack of the particular disulfide bond occurring within EL3 in the $A_{2A}AR$.

With regard to the intracellular loops, the use of the T4L fusion protein strategy in the crystallization process accounts for the highly unstructured nature of IL3 in the GPCRs.¹⁴ This is in good agreement with the rmsf results shown in Figure 1, in

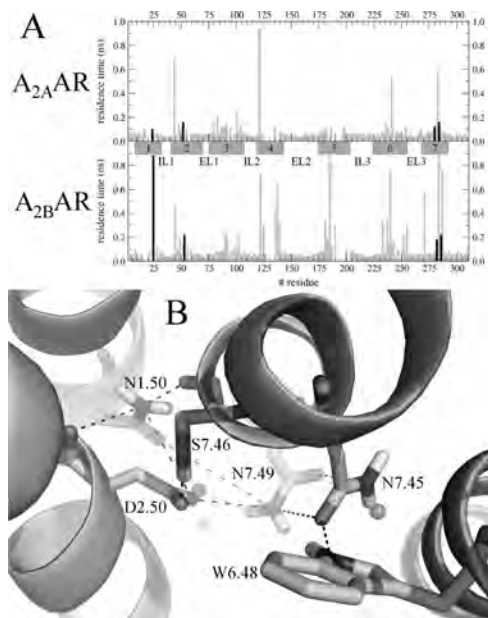


Figure 2. Water analysis of the trajectories of A₂ARs. (A) The residence time of water molecules is represented as a function of the receptor sequence (top, A_{2A}AR; bottom, A_{2B}AR). Residence time was estimated as defined in ref 72. Residues Asn1.50, Asp2.50, Asn7.45, and Asn7.49, interacting with water molecules in the crystal structure of the A_{2A}AR, are highlighted with black bars. Dark gray boxes indicate the location of the TM helices. (B) Insight into the hydrogen bond network connecting helices TM1, TM2, TM6, and TM7 based on the crystal structure of the A_{2A}AR. The residues involved are labeled and shown as sticks, with potential hydrogen bonds depicted as dashed lines and crystallographic waters as spheres. The corresponding hydrogen bond frequencies are listed in Table 3.

particular for the A_{2A}AR, which presents the longest IL3 segment. In contrast, the relatively low rmsf of IL2 agrees with the presence of a α -helical secondary structure in this loop, and the stabilizing interactions between Tyr3.60 in IL3 and the DRY motif in the intracellular tip of TM3. Such behavior was previously observed in MD studies of β 1- and β 2-adrenergic receptors.²⁷

Concerted motions of different domains were identified by principal component analysis (PCA) independently performed on each trajectory (see the Supplementary Methods of the Supporting Information). The extreme structures of the most representative collective motion are shown in Figure 1C, illustrating again the weight of the unstructured loop movement in the protein dynamics. Unexpectedly, the extreme conformations show that, mainly for the A_{2A}AR, TM1 and TM7 slightly separate from each other during the MD trajectories. As discussed below, this movement could be directly related to the disruption of the interaction, initially present in the crystal structure of the A_{2A}AR, between the highly conserved residues in ARs, Glu1.39 and His7.43.

Interactions with the Solvent. The previously presented set of MD simulations avoids the a priori consideration of any water

Table 3. Hydrogen Bond Frequencies in the Structural Motifs Depicted in Figures 2, 5, and 6

hydrogen bond ^a	A _{2A}			A _{2B}	
	R1	R2	RW1	R1	R2
TM1–TM2–TM6–TM7 ^b (Figure 2)					
N1.50–S7.46	81.91%	70.88%	67.93%	70.67%	42.95%
N1.50–D2.50	0.64%	0.84%	3.17%	6.79%	2.86%
D2.50–S7.46	72.96%	99.53%	99.04%	59.05%	48.40%
D2.50–N7.49	20.66%	95.06%	66.18%	68.22%	90.26%
N7.45–N7.49	3.15%	4.33%	0.16%	6.68%	0.00%
N7.45–W6.48	0.00%	0.00%	0.03%	0.00%	0.00%
N1.50–N7.49	0.29%	0.30%	0.00%	0.85%	4.58%
average pairs ^c	1.80	2.71	2.37	2.12	1.89
Ionic Lock ^d (Figure 5)					
D3.49–R3.50 ^e	100%	100%	99.99%	97.31%	99.80%
E6.30–R3.50 ^e	70.71%	99.38%	81.51%	99.38%	98.86%
D3.49–Y3.60	99.10%	81.50%	99.81%	99.88%	99.92%
D3.49–T2.39	93.06%	99.36%	99.66%	61.25%	99.88%
average pairs ^c	3.63	3.80	3.81	3.58	3.98
Toggle Switch ^f (Figure 6)					
H6.52–N5.42	8.78%	1.25%	38.80%	12.11%	40.61%
H6.52–N6.55	4.11%	0.01%	0.19%	1.43%	2.20%
Q3.37–C5.46	94.22%	83.56%	24.52%	79.96%	95.46%
Q3.37–N5.42	4.36%	0.01%	3.11%	62.01%	12.59%
average pairs ^c	1.11	0.85	0.67	1.56	1.51

^a Hydrogen bonding frequency computed with the `g_hbond` utility in GROMACS, using the default parameters (cutoff acceptor–donor–hydrogen angle of 30°, cutoff acceptor–donor radius of 0.35 nm). ^b See Figure 2 for a representation of the polar interactions depicted here. ^c The average number of simultaneous hydrogen bonds per snapshot in the given motif along the MD trajectory. ^d See Figure 5 for a representation of the polar interactions depicted here. ^e See Figure 6 for a representation of the polar interactions depicted here. ^f The occurrence of the salt bridge was analyzed as a regular hydrogen bond.

molecule derived, e.g., from crystallographic observations. This was done to allow a direct comparison between the crystallographic structure of the A_{2A}AR and the homology-derived model of the A_{2B}AR, lacking such information, as it is common in most of the GPCRs of pharmacological interest. Nevertheless, a detailed analysis of the interaction of water molecules with the receptor was performed to ensure a correct behavior of the system, because water molecules have been assigned a role in the structure of GPCRs, and even in the receptor activation process.^{69–71} As a first approach, the maximal residence times in the first hydration shell of every residue (i.e., water molecules a maximum distance of 0.4 nm from the given residue) were estimated as defined by Freites et al.⁷² (see Figure 2A). Remarkably, the cluster of residues connecting TM1, TM2, and TM7, which interact with crystallographic water molecules in the hA_{2A}AR structure (see Figure 2B and next section), show residence times higher than the average in the two receptors, indicating that the absence of such crystal water molecules in the starting structure was successfully restored during the equilibration phase. On the other hand, most of the longest residence times correspond to residues on the intra- or extracellular edges of the helices, where water molecules are not likely to play a structural role.

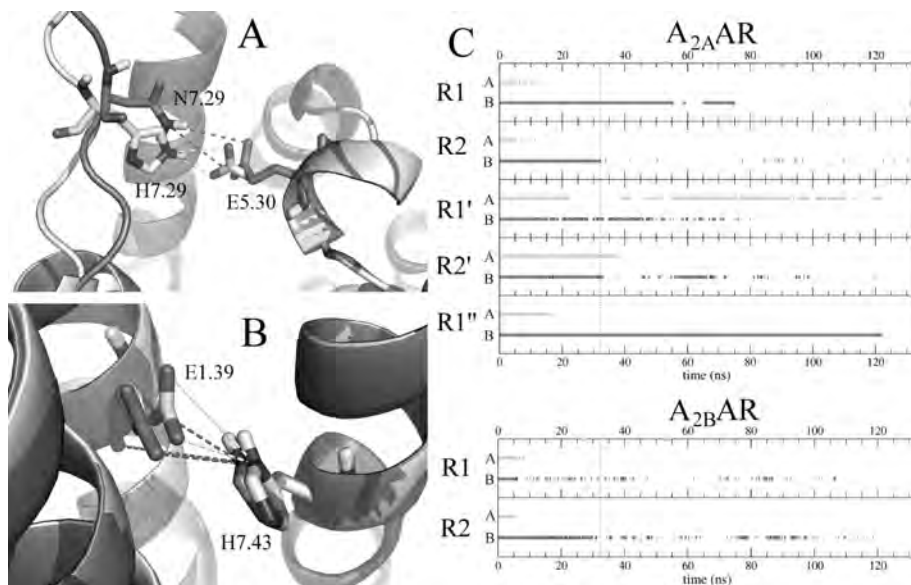


Figure 3. Interactions of specific residues of ARs. (A) Initial interaction between Glu5.30 in EL2 and the residue in position 7.29 in EL3: His²⁶⁴ in A_{2A}AR (white) or Asn²⁶⁶ in A_{2B}AR (gray). (B) Initial interaction between the side chains of Glu1.39 and His7.43 based on the crystal structure of the A_{2A}AR (white) and the homology-derived model of the A_{2B}AR (gray). (C) Hydrogen bonds between Glu5.30 and His- or Asn7.29 depicted in panel A (gray bars) and between Glu1.39 and His7.43 as depicted in panel B (black bars). Results for whole trajectories are shown; the vertical dotted line separates equilibration from the production phase. In the A_{2A}AR, R1 and R2 refer to replicas of the receptor with His²⁶⁴(7.29) in neutral form; R1' and R2' are receptors with His²⁶⁴(7.29) protonated, and R1'' presents both His²⁶⁴(7.29) and His7.43 in TM7 protonated. In the case of charged histidines, both N δ 1 and N ϵ 2 atoms are considered for the computation of hydrogen bonds with the counterpart residue.

To further confirm the feasibility the current protocol ignoring the crystallographic water molecules, we provide an additional MD simulation of the A_{2A}AR, including selected crystallographic water molecules [RW1 (see Table 2)]. The results are summarized in Table 3 and in Figures S1 and S2 of the Supporting Information. We can appreciate that the dynamic properties and the behavior of all structural events analyzed in this work do not significantly deviate from the behavior observed in the other two MD replica simulations of the A_{2A}AR. The frequency of hydrogen bonds of residues connecting TM1, TM2, TM6, and TM7 is particularly conserved, being depicted in Table 3 and discussed in more detail in the next section. A supplementary analysis of all the residues showing polar contacts with the crystallographic water molecules in the A_{2A}AR points in the same direction. The bar graphs in Figure S1 of the Supporting Information represent the number of times that a residue has been ligated to a water molecule for at least 4 ns along the MD trajectories, showing the weak effect that the inclusion of the crystallographic water molecules in the equilibration protocol has on the hydration of the affected residues.

The GPCR-Conserved Hydrogen Bond Network among TM1, TM2, TM6, and TM7. In the crystal structure of the hA_{2A}AR, transmembrane segments TM1, TM2, TM6, and TM7 interact through a GPCR-conserved hydrogen bond network (Figure 2B), which is putatively important in maintaining the architecture of the heptahelical bundle and for the activation mechanism of GPCRs.^{75,74} This network is formed by the following elements, where the percentage of conservation in

the GPCR class-A family³⁸ is indicated in parentheses. The side chain of Asn1.50 (100% conserved) interacts with the backbone of Ser7.46 (63%), the side chain of which interacts with Asp2.50 (94%). A water molecule mediates hydrogen bond interactions among Asp2.50, Asn1.50, and Asn7.49 (75%). Finally, a second water molecule connects the last residue with Asn7.45 (67%), which at the same time is directly hydrogen bonding to Trp6.48 (71%), the so-called toggle switch residue.

The analysis of such a hydrogen bond network, summarized in Table 3, shows the stability of the interactions among residues Asn1.50, Ser7.46, Asp2.50, and Asn7.49. All the involved residue pair hydrogen bonds register high frequencies in R1, R2 replicas of both receptors, and in simulation RW1, with an average number of two simultaneous hydrogen bonds. Interestingly, the complexity of this hydrogen bond network was recently revealed by the recent exploration of the protonation state of Asp2.50 in the β -adrenergic receptors⁷⁵ or by the influence of sodium ions in a recent MD simulation of a model of the D₂ dopamine receptor.⁷⁶

It is worth attending to the time evolution of residue Asn7.45. In the crystal structure of the A_{2A}AR, the oxygen in the side chain of this asparagine is hydrogen bonding to the side chain of Trp6.48. This interaction had been predicted by structural comparison of GPCR models with the crystal structure of rhodopsin, in which the lack of Asn at position 7.45 is supposed to be counterbalanced by crystallographic water molecules mediating interaction with Trp6.48, which is this way arrested

in the *gauche+* (*g+*) rotamer.⁷⁰ However, in the crystal structures of the β -adrenergic receptors, the rotamer of Asn7.45 appears to be flipped with respect to the crystallographic position of this residue in the A_{2A} AR, thus avoiding such interaction with Trp6.48. Our MD simulations of both ARs are in agreement with this last experimental observation, because a change in the rotamer of Asn7.45 has already been registered in the equilibration process of all five simulations (Figure S3 of the Supporting Information), leading to the disruption of the hydrogen bond postulated in the crystal structure of the A_{2A} AR.

Interactions between Specific Residues of the Adenosine Receptor Family. Besides residues conserved in the rhodopsin-A class of the GPCR superfamily, some particular positions of the AR family members deserve special attention, to elucidate their potential structural role.

One particular interaction of the A_{2A} AR crystal structure connects extracellular loops EL2 and EL3 (Figure 3A). In this structure, where the receptor is in complex with the antagonist ZM241385, Glu5.30 (EL2) accepts two hydrogen bonds: one coming from the amino group in the ligand and another coming from His²⁶⁴ (position 7.29) in EL3 making this loop act as a lid that closes the binding site stabilizing the complex (see Figure 4A). Such a role of the ELs in the ligand entrance has been hypothesized in recent molecular modeling of ARs.⁷⁷ In the A_{2B} AR model presented here, the role of this histidine is played by Asn²⁶⁶ (see the multiple-sequence alignment of EL3 of the human ARs in Figure S4 of the Supporting Information), which seemed to point out that a polar interaction, rather than a salt bridge, would be conserved between the two loops. This possibility was not discarded by the pK_a prediction of PDB2PQR, which assigns the neutral form for His7.29 (pK_a = 7.03), which we subsequently considered in the first instance for our simulations (replicas R1 and R2 of the A_{2A} AR). However, such interaction is broken in the equilibration phase of two replicas simulated per receptor (Figure 3C), a result that would lead to the hypothesis that in the apo form of the receptor, this interaction is easily broken to allow access of extracellular diffusible molecules to the binding site. On the other hand, further ionization state calculations with MCCE predicted His7.29 with a net positive charge, which encouraged us to simulate two independent replicas of the apo state of the A_{2A} AR considering this residue charged. As one can see in Figure 3C (R1' and R2'), the interaction initially maintained between EL2 and EL3 through a salt bridge between His7.29 and Glu5.30 is frequently observed (61% of the time) in R1', while in R2' the frequency of this interaction is only slightly higher (6%) than in the cases where His7.29 was modeled neutral (0% occurrence). Because it is not clear from the simulations whether in the apo state this histidine is positively charged, the next step was to investigate whether a charged histidine could favor the closed state of the extracellular loops as observed in the crystal structure of the hA_{2A} AR in complex with an antagonist (Figure 4A). Subsequent MD simulations of the A_{2A} AR in the presence of the ligand were performed. Here, a neutral His7.29 in EL2 [simulation RL1 (Table 2)] promotes an opening of the lid formed by EL2 and EL3, resulting in the loss of most receptor–ligand interactions (see Figure 4B, top panel). On the other hand, a charged histidine in that position [simulation RL1' (Table 2)] maintains such a lid atop the ligand and the complex remains stable for the rest of the simulation (Figure 4B, bottom panel), thus pointing out a preference for the protonated state of this histidine in the holo form of the receptor.

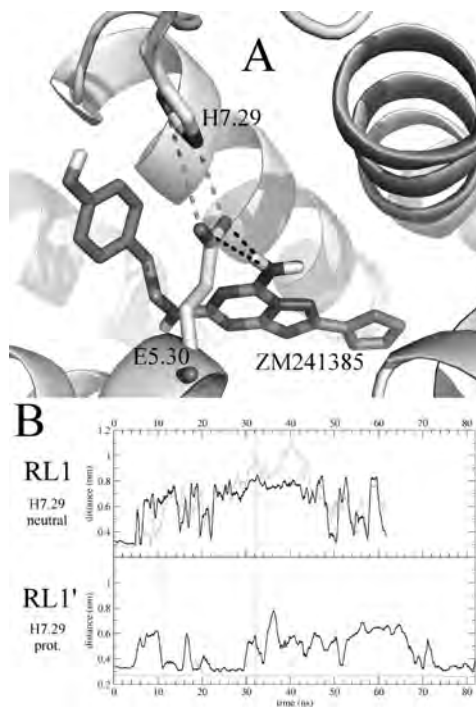


Figure 4. Influence of the protonation state of His²⁶⁴(7.29) in the dynamics of the crystallographic protein–ligand complex. (A) Representation of the initial structure of the A_{2A} AR (white) in complex with ZM241385 (gray). Key residues are labeled. (B) Measurement of the minimum distances between O ϵ atoms of Glu5.30 and (i) the nitrogen of the ligand's exocyclic amino group (black lines) and (ii) the nitrogen of His7.29 (gray lines), as depicted by dashed lines in panel A. Results of whole trajectories are shown, with vertical dotted lines indicating the release of the positional restraints for the ligand (11 ns) and the release of restraints for the α -carbon atoms in the loops (32 ns).

We also focused on two residues completely conserved in the ARs family: a glutamic acid in position 1.39 and a histidine in position 7.43, which are close in space in the A_{2A} AR crystallographic structure and had been previously postulated to interact through a hydrogen bond.^{78–80} To initially form such a hydrogen bond (Figure 3B), the histidine residue was protonated at N δ , as suggested by several of the considered prediction tools (Schrödinger, Molprobit, and PDB2PQR), even though the geometry of the hydrogen bond in the crystal structure is not associated with a strong interaction, taking into account the ideal geometries derived from the study of the Baker group⁸¹ [θ angle = 99°; ψ angle = 88°; δ_{HA} = 0.25 nm (see the original reference for details of these values)]. Accordingly, MD simulations with His7.43 in its neutral form show that this hydrogen bond tends to be transient or even completely broken, only being significantly occupied (27%) in R1 of the A_{2A} AR, but with an occupancy of <8% in the remaining simulations, including both replicas of the A_{2B} AR (Figure 3C). It is worth noting that His7.43 has been experimentally involved

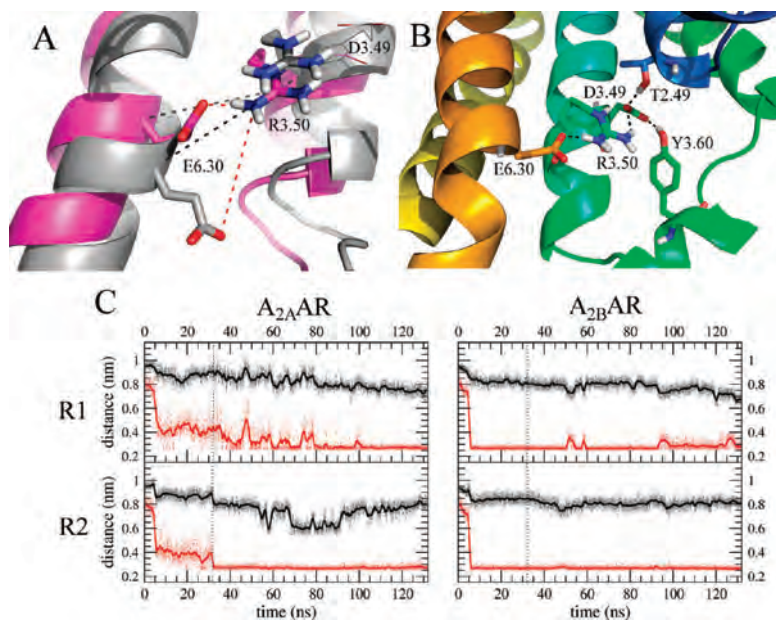


Figure 5. Analysis of the ionic lock. (A) Salt bridge between Arg3.50 and Glu6.30, commonly termed the ionic lock, in the A_{2A}AR. The initial configuration (gray) and a stable snapshot (magenta) are superimposed. (B) Representative snapshot of the hydrogen bond network in which the ionic lock is contextualized. The residues involved in helices TM2 (cyan), TM3 (green), and TM6 (orange) are labeled and shown as sticks, with potential hydrogen bonds depicted as dashed lines. The corresponding hydrogen bond frequencies are listed in Table 3. (C) For R1 and R2 of both receptors, the time evolution of the minimum distance between any O ϵ atom of Glu6.30 and any N ζ atom of Arg3.50 is colored red (thick line representing the smoothed version). Similarly, the corresponding distance between the α -carbons of the same residues is represented in black. Results for whole trajectories are shown; the vertical dotted line separates equilibration from the production phase.

in agonist binding,⁸² presumably through interaction with the hydroxyl groups of the ribose ring of those ligands. It remains unclear whether the proposed agonist–receptor interaction should occur through hydrogen bond acceptance of N ϵ of His7.43 (meanwhile, its N δ interacts with Glu1.39⁸⁰) or whether this interaction with the agonists promotes a disruption of the hydrogen bond with the adjacent Glu1.39. In Figure 3C, the fact that this interaction is spontaneously broken in the absence of any ligand can be appreciated, thus suggesting an important role of this residue in the conformational equilibrium of the receptor,⁷⁸ probably connected with the structural changes described above for TM7 (Figure 1C). Conversely, a salt bridge interaction is strongly suggested by the pK_a calculations performed with MCCE on this pair of residues. This possibility was considered in an independent MD simulation of the A_{2A}AR, where both His7.29 and His7.43 are modeled in their charged form, from now on simulation R1'. Figure 3C shows how the established salt bridge of a positively charged His7.43 with Glu1.39 is sufficiently strong to maintain a tight interaction between TM1 and TM7. Accordingly, the separation of TM1 and TM7 is not observed in a PCA analysis of this replica (Figure S5 of the Supporting Information), as opposed to what is shown in Figure 1 when the His7.43 is modeled in its neutral form. Alternate protonation states, such as a protonated form of Glu1.39, were not suggested by any of the pK_a prediction software used.

Ionic Lock. The TM3–TM6 ionic lock, linking the (D/E)RY motif in TM3 with Glu6.30 in TM6, is broken in the crystal structure of the A_{2A}AR, which happened in the β 1- and β 2-adrenergic receptor crystal structures and in contrast to what was expected in an inactive form of a GPCR.²³ As one can appreciate in Figure 5, the ionic lock is achieved in main simulations performed on both A_{2A}AR and A_{2B}AR receptors (replicas R1 and R2). In the A_{2A}AR, this interaction is formed at the end of the equilibration phase (R2) or even after 50 ns of the production phase (R1), while in both replicas of the A_{2B}AR, the lock is formed just after a few nanoseconds of equilibration time. In all four cases, once the lock is formed it can be considered strong and stable for the rest of the simulation time (see Table 3). It is also important to note that in all these cases TM3 and TM6 approach each other, as measured by a decrease in the distance between the α -carbons of Glu6.30 and Arg3.50 ($C\alpha$ – $C\alpha$ distance, comparing the initial value with the averaged distance of the last 20 ns, reduced in a magnitude by 0.137 to 0.232 nm, depending on the replica). The whole network of polar interactions is depicted in Figure 5B and quantified in Table 3. Such interactions involve not only salt bridges between Arg3.50 and both Asp3.49 and Glu6.30 but also stable hydrogen bonds between Asp3.49 and Tyr3.60 in IL2 and Thr2.39 in TM2. This particular network around Asp3.49, with almost four simultaneous interactions as an average, might be responsible of the strong salt

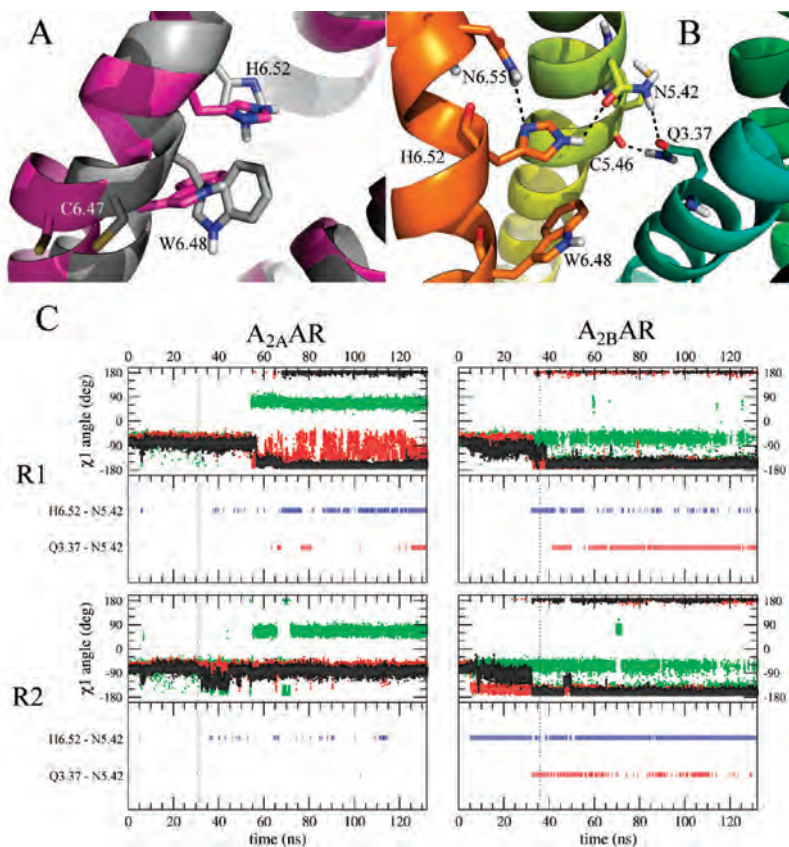


Figure 6. Analysis of the toggle switch. (A) Initial phase (gray) and a stable snapshot of the production phase (magenta) of the side chain configuration of residues Cys6.47, Trp6.48, and His6.52, showing the formation of the so-called toggle switch in the A_{2A} AR. (B) Three-dimensional representation of the potential hydrogen bonds (dashed lines) between the residues surrounding the toggle switch (sticks), located at helices TM3 (cyan), TM5 (yellow), and TM6 (orange). The corresponding hydrogen bond frequencies are listed in Table 3. (C) A separate panel for each MD simulation represents the time evolution of the χ_1 dihedral angles of the residues involved in the toggle switch (top part): Trp6.48 (black dots), His6.52 (red dots), and Cys6.47 (green dots). The bottom part of each panel represents the evolution of the hydrogen bonds of Asn5.42 with His6.52 (blue bars) or with Gln3.37 (red bars). Results of whole trajectories are shown; the vertical dotted line separates equilibration from the production phase.

bridge of this residue with the adjacent Arg3.50, which has been proposed as the primary constituent of the ionic lock by Vogel et al.²¹

Conformational Changes Accompanying the Toggle Switch. The conformational changes associated with the GPCR conserved motif, (6.47)CWXP(6.50), have been thoroughly examined in the A_{2A} AR and A_{2B} AR structures (Figure 6). The χ_1 dihedral angles of selected residues were recorded along the different MD simulations. The starting rotameric state of Trp6.48 in both receptors corresponds to the g^+ rotamer ($\chi_1 \sim -80^\circ$), which is associated with the inactive conformation. On the other hand, the toggle switch is considered active when this residue adopts a *trans* (*t*) conformation; i.e., the χ_1 dihedral is in the equivalent regions defined between -120° and -180° or between 120° and 180° .²² This rotameric transition might be

concerted with conformational changes in positions 6.52 and 6.47, as described in the β_2 -adrenergic receptor.²² Position 6.52 is frequently an aromatic residue in class-A GPCRs (82% Phe), while position 6.47 is a Cys in 74% of the cases.³⁸ Whereas the last residue is conserved in the two A_2 AR subtypes, the former is occupied by a histidine in all ARs members, with the exception being the A_3 AR, where a serine is present. In the crystal structure of the A_{2A} AR, both His6.52 and Cys6.47 show approximately the same χ_1 angle as residue Trp6.48 (i.e., g^+ conformation). As depicted in Figure 6, we observe the transition of the toggle switch in R1 of the A_{2A} AR (after 24 ns in the production phase) and in the two replicas of the A_{2B} AR (where the change occurs just after the equilibration phase). In all four cases, there is a clear correlation between the rotameric states of the toggle switch (Trp6.48) and residue His6.52, because the last also suffers a

transition from the *g*+ rotamer to the *t* rotamer only when the toggle switch changes its conformation. Remarkably, in R2 of the A_{2B} AR, the conformational change in His6.52 clearly precedes the formation of the toggle switch, suggesting that the rotameric change of the aromatic residue in position 6.52 (His in the ARs considered here) might favor the formation of the toggle switch. In R1 of the A_{2A} AR, the *t* conformation of His6.52 prevails 59% of the time after the irreversible formation of the toggle switch, which occurs simultaneously with the conformational change of His6.52. On the other hand, the toggle switch does not occur in R2 of the A_{2A} AR along the simulated time, and consequently, the rotamer of His6.52 stays in the native *g*+ state. Finally, we have investigated the possible correlation between the toggle switch and the rotameric state of Cys6.47, as previously suggested on the basis of molecular modeling of the β_2 -adrenergic receptor.²² We observe that, in the A_{2A} AR simulations, Cys6.47 changes its rotameric state in both R1 and R2 simulations, while the toggle switch occurs in only R1. On the other hand, Cys6.47 does not change its conformation in any of the two simulations of the A_{2B} AR, where the toggle switch invariably occurs. Thus, our simulations do not support a correlation of the rotameric state of Cys6.47 with the formation of the toggle switch.

It is important to note the interactions of His6.52 with Asn5.42 in TMS through a side chain hydrogen bond, which is somehow concerted with the rotameric state of the former residue in TM6. The complete hydrogen bond network, which also involves residues in TM3, is depicted in Figure 6B and quantitatively analyzed in Table 3. It can be appreciated in Figure 6C that His6.52 needs to adopt the *t* rotameric state to form the hydrogen bond with Asn5.42. As a consequence, the frequency of this hydrogen bond is significant only in the three cases when the toggle switch is formed (i.e., all cases except R2 in the A_{2A} AR), although only in one case (R2 of the A_{2B} AR) can this interaction be considered stable. It is worth noting that Asn6.55, which is crucial for the ligand binding,¹² does not form any stable interaction with the residues in the vicinity in the MD simulations (performed in the absence of any ligand). In particular, the possibility of an interaction with His6.52, located almost one α -helix turn below, is not feasible according to the data presented in Table 3. On the other hand, Asn5.42 tends to form a hydrogen bond with Gln3.37 once the toggle switch is formed. Gln3.37 also strongly connects to TMS through a hydrogen bond with the main chain of Cys5.46 (Figure 6B and Table 3). The consideration of the crystallographic water molecules on the A_{2A} AR MD simulation (RW1, with W565 in the vicinity of the aforementioned residues) results in only subtle effects on the identified network: the interaction between TM6 and TMS (His6.52 and Asn5.42) is more prominent, whereas the interaction between TMS and TM3 (Cys5.46–Gln3.37) is somehow weakened as compared to the MD simulations without any crystallographic waters.

DISCUSSION

The release of the hA_{2A} AR structure in an inactive conformation offers an excellent starting point for a deep exploration of the conformational and dynamic properties of the adenosine receptor family. In this work, we have studied the two human A_2 receptors (hA_{2A} AR and hA_{2B} AR) through all-atom MD simulations, generating independent replicas for each receptor. This schema allows extraction of more robust conclusions than if only one replica per receptor were performed, even if the duration of

the simulation would be longer. Overall, a general analysis of the MD trajectories shows good agreement with the expected behavior of a GPCR, with higher mobility of the loop regions, while TM segments remain more stable. Interestingly, the main differences between the general dynamic behaviors of the A_{2A} AR and A_{2B} AR explored can be explained in terms of their particular sequences, loops lengths, and disulfide bridges. The observed enhanced mobility of the extracellular part of the A_{2B} AR could justify, to some extent, the lower affinity of this receptor for the natural ligand, adenosine,⁸³ as a function of the expected higher rate of diffusion of ligand molecules in the extracellular side. Accordingly, the consideration of a positively charged histidine in EL3 of the hA_{2A} AR also reduces the mobility of the extracellular regions of this receptor. The conservation pattern shown in Figure S4 of the Supporting Information denotes that this histidine is only present in the A_1 AR and A_{2A} AR, which are precisely the receptors in the family showing the highest affinity for adenosine. These observations reinforce the idea that the different composition of the extracellular region surrounding the binding site is modulating ligand affinity, recently proposed by us in the design of hA_3 AR ligands,⁸⁴ as well as by other authors on the basis of sequence analysis and site-directed mutagenesis studies.⁸⁵ Interestingly, the crystal structure of the A_{2A} AR in complex with an agonist, which has been released while this paper was being revised,⁸⁶ shows a disruption of the interactions between EL2 and EL3 to accommodate the bulky substitution at the N6 exocyclic amino group of the cocrystallized ligand. We thus find this experimental result in good agreement with the simulations reported herein, in which the lid formed by these loops tends to open in the absence of any ligand.

We also investigated the structural role of another relevant histidine located at position 7.43 in all ARs. The behavior of the interaction between the Glu1.39–His7.43 conserved pair is somehow surprising. Both in the apo form and in complex with the antagonist, this interaction fluctuates in the best cases or is even broken in some replicas. Conversely, when the charged form of this histidine is considered, the Glu1.39–His7.43 ionic pair is maintained along the whole MD simulation (see Figure 3C, simulation R1'). These results should be contextualized with the mutagenesis data available for this particular pair of residues. In the A_{2A} AR, a mutation of His7.43 to either Asp or Glu does not produce any detectable changes in either agonist or antagonist binding.⁸⁷ In this case, pK_a calculations predict that at least one of the two interacting residues should be modeled in its neutral, protonated form (see Figure S6 of the Supporting Information), which could thus be used to argue against the possibility of a salt bridge interaction between Glu1.39 and His7.43 on the wild-type receptor. However, a polar interaction between side chains at these positions may be a minimum condition for the integrity of the binding site, as a His7.43Ala mutant yields no detectable agonist or antagonist binding.⁸⁸ Finally, it should be noted that a Glu1.39Gln mutant has little effect on agonist binding, which is more affected by a His7.43Tyr mutation,⁷⁸ which led to the idea of position 7.43 interacting with the ribose moiety,⁷⁹ recently confirmed by the crystal structure of the agonist-bound A_{2A} AR.⁸⁶ Altogether, these data support a polar interaction between the aforementioned positions in TM1 and TM7, which is presented as a weak, intermittent interaction according to our MD simulations (Figure 3C, simulations R1, R2, R1', and R2'), allowing a gradual separation of TM1 and TM7 that could be related to the conformational equilibrium of the receptor. Alternatively, the hypothesis of a salt bridge

occurring between these residues implies that no separation between these helices occurs along the simulation time, precluding a dynamic role of this pair of residues in the modulation of receptor conformations. The reliability of this alternate hypothesis should be tested with the expression and pharmacological characterization of a mutant receptor where His7.43 is replaced with a positively charged residue (Lys), as modeled in Figure S6 of the Supporting Information.

Several important dynamic properties, which prompt novel receptor interactions not observed in the crystal structure of the inactive A_{2A} AR, are reported and related to the conformational equilibrium of the ARs. First, the ionic lock, linking the (D/E)RY motif in TM3 with Glu6.30 in TM6, which is absent in the starting receptor structures, is spontaneously formed in the MD simulations presented. The fact that the ionic lock was also broken in the crystal structures of the β -adrenergic receptors has received much attention in the recent computational studies of this receptor family. The conclusions extracted from independent microsecond MD simulations^{26,27} suggest that either the removal of T4 lysozyme or the absence of the G protein in the crystal structures might favor the formation of the ionic lock. The observation of this event in the initial steps of our MD simulations of the ARs, and the subsequent approach of the cytoplasmic part of the TM3 and TM6 helices, are in agreement with this hypothesis and with recently reported MD simulations of the A_{2A} AR,³⁶ while other authors have noted that such a conformational change is observed only at high saline concentrations in the same receptor.³⁰ Our results support the idea of a more complex network of hydrogen bonds and salt bridges stabilizing the ionic lock, where all four polar interactions considered among residues Thr2.39, Asp3.49, Tyr3.60, Arg3.50, and Glu6.30 are quite stable in our simulations (see Table 3). Interestingly, the importance of the salt bridge between the conserved Asp3.49 and Arg3.50 [within the (D/E)RY motif in TM3] has been identified with infrared spectroscopy experiments in rhodopsin²¹ and was supported by previous experiments with the β_2 receptor.²⁰

A second event that we have identified is the spontaneous formation of the toggle switch in TM6, a closer look at which reveals novel structural elements in the ARs that deserve further analysis. In particular, the rotation of the side chain of Trp6.48 is accompanied in our simulations by a concerted rotameric transition of His6.52, a typical residue of the AR family with the only exception being the A_3 AR. Interestingly, a histidine is also present at this position on the ghrelin receptor, where its key role in the basal activity of that receptor was postulated and experimentally verified.⁸⁹ Moreover, our simulations have revealed additional contacts of His6.52 with Phe5.43 and Asn5.42. Available mutagenesis data on the A_{2A} AR indicate that mutation of any of these three residues by an alanine precludes the detection of ligand binding,⁸⁸ which is partially reverted by the restoration of aromatic side chains in position 6.52 or 5.43 or a polar amino acid (Ser) at position 5.42. The phenylalanine at position 5.43 plays a role in the packing between TMS and TM6, but we did not observe any conformational change for this residue in any of the independent MD simulations considered. This result is in contrast with the conclusions extracted from the recent MD simulations of the hA_{2A} AR by Lyman et al.,³⁶ who suggested that the occurrence of the toggle switch should be associated with conformational changes in residue Phe¹⁸² (5.43). In fact, such an association was motivated by the assignment of this residue to position 5.47 by the authors, and not to actual position 5.43, thus being incorrectly aligned with the

Phe5.47 that is part of the highly conserved aromatic cluster of the aminergic receptor family.⁹⁰ Remarkably, residues at both positions 5.47 and 6.52 are conserved in the A_1 AR, A_{2A} AR, and A_{2B} AR, with a small valine at position 5.47 favoring the rotameric transition of His6.52. Conversely, a correlated mutation is observed in all A_3 ARs, where the histidine in position 6.52 is substituted with a Ser while a more bulky Ile substitutes for the otherwise subtype-conserved Val5.47. This evolutionary mechanism of a correlated mutation in positions 5.47 and 6.52 further reinforces the idea of a structural role of this pair of positions. The network of interactions that stabilize this microenvironment of the toggle switch in the A_2 ARs reveals a previously unobserved dynamic role of polar residues Asn5.42 and Gln3.37. Our simulations indicate that the new connections established among transmembrane helices TM3, TMS, and TM6 are crucial for the receptor architecture or even the activation process, which might provide further explanation for the site-directed mutagenesis available for Gln3.37^{91,92} and Asn5.42,⁸⁸ in contrast with the hypothesis formulated before the release of the A_{2A} AR crystallographic structure that had connected Gln3.37 with ligand binding.⁸⁷ This result is in agreement with the activation mechanism proposed in light of the new structures of agonist-bound β -adrenoreceptors, released during the preparation of this work.^{17,18} In that family of GPCRs, the main variation in the experimental conformation of the agonist-bound receptor with respect to previous antagonist-bound structures is the new interaction between residues at positions 5.43 and 6.55 (one helix turn above His6.52), bringing TMS and TM6 closer by $\sim 1 \text{ \AA}$ ¹⁷ or even 2 \AA .¹⁸ However, the fully activated structures of the β_2 -adrenoreceptor,¹⁹ and opsin,¹⁶ as well as the very recent crystal structure of the agonist-bound A_{2A} AR,⁸⁶ show no change in the side chain conformation of the highly conserved Trp6.48 (toggle switch). A recent hypothesis by the Schwartz group suggests that the switch of Trp6.48, which is only 70% conserved in the GPCR family, would be one of the many potential microswitches occurring during GPCR activation and thus would be dispensable in a given GPCR (i.e., β_2 -adrenoreceptor), which might still conserve a tryptophan in this position.⁹³ In other GPCRs, residue Trp6.48 would change the rotameric state in the initial steps of the activation process, as previously postulated by several authors,^{22,90} preceding the higher-magnitude conformational changes observed in TM6.^{18,19} In light of the new experimental evidence showing an unchanged rotameric state of Trp6.48 between the agonist- and antagonist-bound conformations of the A_{2A} AR, we hypothesize that the role of this switch is transient, promoting an intermediate conformation that would favor the achievement of major movements observed on TM6 in the active-like conformation of the A_{2A} AR. In fact, a closer look at the major conformational changes between the two crystal structures of this receptor reveals that the movement of the intracellular side of TM6 is accompanied by the rotameric transition of aromatic residues Phe5.62 and Tyr5.68, in the intracellular side of TMS.⁸⁶ To avoid steric clashes of these moving residues with TM6, this helix must visit a more open conformation different from that observed in either antagonist- or agonist-bound crystal structures, a conformation that might be promoted by the toggle switch and surrounding interaction network described in this work. Notably, the superposition of the end point conformation of the A_{2A} AR R1 simulation with the agonist-bound crystal structure shows a remarkable overlay of TM6 (see Figure S7 of the Supporting Information), which in both cases is shifted away from the TM bundle on its intracellular

side (that is, below the conserved proline kink). The observations reported herein suggest that further site-directed mutagenesis experiments with the aforementioned positions in helices TMS, TM6, and TM7, especially considering mutations less drastic than those previously reported, would shed light on the molecular mechanisms of AR activation.

■ ASSOCIATED CONTENT

5 Supporting Information. Supplementary Methods detailing trajectory analyses; structure of the A_{2A}AR replica with crystallographic waters (RW1), with a water occupancy analysis of the first hydration shell for the residues in the vicinity of the crystallographic waters (Figure S1); rmsd analysis of loop equilibration and production trajectories, and comparative rmsf analysis of RW1 (Figure S2); analysis of the conformational states of Asn7.45 (Figure S3); multiple-sequence alignment of EL3 of human ARs (Figure S4); PCA of the R1'' simulation (Figure S5); structural models and short MD simulations of mutant receptors for position 7.43 (Figure S6); and comparison of starting and end point structures belonging to the R1 simulation with the crystal structure of the agonist-bound A_{2A}AR (Figure S7). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hugo.teran@usc.es. Phone: +34 881813873. Fax: +34981951473.

Funding Sources

Funding from the Xunta de Galicia, Consellería de Sanidade (Grant PS09/63), and the Spanish Ministry of Science and Innovation (2010 ICTS Project) is gratefully acknowledged. Á.P. and H.G.-d.T. are researchers of the Isidro Parga Pondal program (Xunta de Galicia, Spain). D.R. is a recipient of a predoctoral grant from the Fondo de Investigación Sanitaria (ISCIII, MICINN).

■ ACKNOWLEDGMENT

Computations were conducted at the resources of the Galician Supercomputing Center (CESGA). We thank Pilar Brocos for assistance in the design of automated protocols. We acknowledge reviewer 1 for important suggestions that contributed to improve this manuscript.

■ ABBREVIATIONS

AR, adenosine receptor; GPCR, G protein-coupled receptor; EL and IL, extracellular and intracellular loops, respectively; MD, molecular dynamics; POPC, palmitoyloleoylphosphatidylcholine; PCA, principal component analysis; rmsd, root-mean-square deviation; rmsf, root-mean-square fluctuation; TM, transmembrane; hA_{2A}AR and hA_{2B}AR, human receptors; A_{2A} and A_{2B}, adenosine receptors; D₃DR, D₃ dopamine receptor; tβ1, turkey β1-adrenergic receptor; hβ2, human β2-adrenergic receptor; g+, *gauge+*; t, *trans*.

■ REFERENCES

(1) Rosenbaum, D. M., Rasmussen, S. G., and Kobilka, B. K. (2009) The structure and function of G-protein-coupled receptors. *Nature* 459, 356–363.

(2) Kenakin, T. (2004) Principles: Receptor theory in pharmacology. *Trends Pharmacol. Sci.* 25, 186–192.

(3) Hanson, M. A., and Stevens, R. C. (2009) Discovery of new GPCR biology: One receptor structure at a time. *Structure* 17, 8–14.

(4) Fredriksson, R., Lagerström, M. C., Lundin, L.-G., and Schiöth, H. B. (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256–1272.

(5) Jacobson, K. A., and Gao, Z.-G. (2006) Adenosine receptors as therapeutic targets. *Nat. Rev. Drug Discovery* 5, 247–264.

(6) Ohta, A., and Sitkovsky, M. (2001) Role of G-protein-coupled adenosine receptors in downregulation of inflammation and protection from tissue damage. *Nature* 414, 916–920.

(7) Kalla, R. V., and Zablocki, J. (2009) Progress in the discovery of selective, high affinity A(2B) adenosine receptor antagonists as clinical candidates. *Purinergic Signalling* 5, 21–29.

(8) Downey, J. M., Liu, G. S., and Thornton, J. D. (1993) Adenosine and the anti-infarct effects of preconditioning. *Cardiovasc. Res.* 27, 3–8.

(9) Miller, D. D. (2005) Impact of selective adenosine A2A receptor agonists on cardiac imaging feeling the lightning, waiting on the thunder. *J. Am. Coll. Cardiol.* 46, 2076–2078.

(10) Fredholm, B. B., Arslan, G., Kull, B., and Svenningsson, P. (1998) Locating the neuronal targets for caffeine. *Drug Dev. Res.* 45, 324–328.

(11) Franco, R., Ferre, S., Torvinen, M., Gines, S., Hillion, J., Ciruela, F., Canela, E. L., Mallol, J., Casado, V., Lluís, C., and Fuxe, K. (2001) Adenosine/dopamine receptor-receptor interactions in the central nervous system. *Drug Dev. Res.* 52, 296–302.

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

(13) Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G. F., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. L., Kobilka, B. K., and Stevens, R. C. (2007) High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* 318, 1258–1265.

(14) Jaakola, V.-P., Prilusky, J., Sussman, J. L., and Goldman, A. (2005) G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng. Des. Sel.* 18, 103–110.

(15) Martinelli, A., and Tuccinardi, T. (2008) Molecular modeling of adenosine receptors: New results and trends. *Med. Res. Rev.* 28, 247–277.

(16) Park, J. H., Scheerer, P., Hofmann, K. P., Choe, H.-W., and Ernst, O. P. (2008) Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* 454, 183–187.

(17) Warne, T., Moukhametzianov, R., Baker, J. G., Nehme, R., Edwards, P. C., Leslie, A. G., Schertler, G. F., and Tate, C. G. (2011) The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* 469, 241–244.

(18) Rosenbaum, D. M., Zhang, C., Lyons, J. A., Holl, R., Aragao, D., Arlow, D. H., Rasmussen, S. G., Choi, H. J., Devree, B. T., Sunahara, R. K., Chae, P. S., Gellman, S. H., Dror, R. O., Shaw, D. E., Weis, W. L., Caffrey, M., Gmeiner, P., and Kobilka, B. K. (2011) Structure and function of an irreversible agonist-β2 adrenoceptor complex. *Nature* 469, 236–240.

(19) Rasmussen, S. G., Choi, H. J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., Devree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzk, I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. L., and Kobilka, B. K. (2011) Structure of a nanobody-stabilized active state of the β2 adrenoceptor. *Nature* 469, 175–180.

(20) Ballesteros, J. A., Jensen, A. D., Liapakis, G., Rasmussen, S. G., Shi, L., Gether, U., and Javitch, J. A. (2001) Activation of the β2-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* 276, 29171–29177.

(21) Vogel, R., Mahalingam, M., Lüdeke, S., Huber, T., Siebert, F., and Sakmar, T. P. (2008) Functional role of the “ionic lock”: An

interhelical hydrogen-bond network in family A heptahelical receptors. *J. Mol. Biol.* 380, 648–655.

(22) Shi, L., Liapakis, G., Xu, R., Guarnieri, F., Ballesteros, J. A., and Javitch, J. A. (2002) β_2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch. *J. Biol. Chem.* 277, 40989–40996.

(23) Audet, M., and Bouvier, M. (2008) Insights into signaling from the β_2 -adrenergic receptor structure. *Nat. Chem. Biol.* 4, 397–403.

(24) Kobilka, B. K., and Deupi, X. (2007) Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* 28, 397–406.

(25) Chien, E. Y. T., Liu, W., Zhao, Q., Katritch, V., Han, G. W., Hanson, M. A., Shi, L., Newman, A. H., Javitch, J. A., Cherezov, V., and Stevens, R. C. (2010) Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 330, 1091–1095.

(26) Vanni, S., Neri, M., Tavernelli, I., and Rothlisberger, U. (2009) Observation of “ionic lock” formation in molecular dynamics simulations of wild-type β_1 and β_2 adrenergic receptors. *Biochemistry* 48, 4789–4797.

(27) Dror, R. O., Arlow, D. H., Borhani, D. W., Jensen, M. Ø., Piana, S., and Shaw, D. E. (2009) Identification of two distinct inactive conformations of the β_2 -adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4689–4694.

(28) Romo, T. D., Grossfield, A., and Pitman, M. C. (2010) Concerted interconversion between ionic lock substates of the β_2 adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* 98, 76–84.

(29) Huber, T., Menon, S., and Sakmar, T. P. (2008) Structural basis for ligand binding and specificity in adrenergic receptors: Implications for GPCR-targeted drug discovery. *Biochemistry* 47, 11013–11023.

(30) Jójárt, B. z., Kiss, R. b., Viskolcz, B. I., Kolosváry, I. n., and Keserü, G. r. M. (2010) Molecular Dynamics Simulation at High Sodium Chloride Concentration: Toward the Inactive Conformation of the Human Adenosine A2A Receptor. *J. Phys. Chem. Lett.* 1, 1008–1013.

(31) Hu, J., Wang, Y., Zhang, X., Lloyd, J. R., Li, J. H., Karpiak, J., Costanzi, S., and Wess, J. (2010) Structural basis of G protein-coupled receptor-G protein interactions. *Nat. Chem. Biol.* 6, 541–548.

(32) Kolb, P., Rosenbaum, D. M., Irwin, J. J., Fung, J. J., Kobilka, B. K., and Shoichet, B. K. (2009) Structure-based discovery of β_2 -adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6843–6848.

(33) Bhattacharya, S., Hall, S. E., and Vaidehi, N. (2008) Agonist-induced conformational changes in bovine rhodopsin: Insight into activation of G-protein-coupled receptors. *J. Mol. Biol.* 382, 539–555.

(34) Provasi, D., and Filizola, M. (2010) Putative active states of a prototypic g-protein-coupled receptor from biased molecular dynamics. *Biophys. J.* 98, 2347–2355.

(35) Carlsson, J., Yoo, L., Gao, Z.-G., Irwin, J. J., Shoichet, B. K., and Jacobson, K. A. (2010) Structure-based discovery of A2A adenosine receptor ligands. *J. Med. Chem.* 53, 3748–3755.

(36) Lyman, E., Higgs, C., Kim, B., Lupyan, D., Shelley, J. C., Farid, R., and Voth, G. A. (2009) A role for a specific cholesterol interaction in stabilizing the Apo configuration of the human A_{2A} adenosine receptor. *Structure* 17, 1660–1668.

(37) Baldwin, J. M., Schertler, G. F., and Unger, V. M. (1997) An α -carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* 272, 144–164.

(38) Mirzadegan, T., Benko, G., Filipek, S., and Palczewski, K. (2003) Sequence analyses of G-protein-coupled receptors: Similarities to rhodopsin. *Biochemistry* 42, 2759–2767.

(39) Mobarec, J. C., Sanchez, R., and Filizola, M. (2009) Modern homology modeling of G-protein coupled receptors: Which structural template to use? *J. Med. Chem.* 52, 5207–5216.

(40) Ballesteros, J. A., and Weinstein, H. (1995) Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. In *Methods in Neuroscience*, pp 366–428, Academic Press, San Diego.

(41) Sali, A., and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779–815.

(42) Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci.* 9, 1753–1773.

(43) Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26, 283–291.

(44) Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35, W375–W383.

(45) Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* 32, W665–W667.

(46) Song, Y., Mao, J., and Gunner, M. R. (2009) MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling. *J. Comput. Chem.* 30, 2231–2247.

(47) Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., and Honig, B. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* 23, 128–137.

(48) Song, Y., Mao, J., and Gunner, M. R. (2003) Calculation of proton transfers in bacteriorhodopsin bR and M intermediates. *Biochemistry* 42, 9875–9888.

(49) *Macromodel* (2009) Schrödinger, LLC, New York.

(50) Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236.

(51) Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105, 6474–6487.

(52) Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.

(53) Michino, M., Abola, E., Brooks, C. L., Dixon, J. S., Moul, J., and Stevens, R. C. (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discovery* 8, 455–463.

(54) Chakrabarti, N., Neale, C., Payandeh, J., Pai, E. F., and Pomès, R. (2010) An iris-like mechanism of pore dilation in the CorA magnesium transport system. *Biophys. J.* 98, 784–792.

(55) Lindahl, E., and Edholm, O. (2000) Mesoscopic undulations and thickness fluctuations in lipid bilayers from molecular dynamics simulations. *Biophys. J.* 79, 426–433.

(56) Berger, O., Edholm, O., and Jähnig, F. (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72, 2002–2013.

(57) Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981) Interaction models for Water in Relation to Protein Hydration. In *Intermolecular Forces* (Pullman, B., Ed.), pp 331–342, D. Reidel Publishing Co., Dordrecht, The Netherlands.

(58) Nose, S., and Klein, M. L. (1983) Constant pressure molecular-dynamics for molecular systems. *Mol. Phys.* 50, 1055–1076.

(59) Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single-crystals: A new molecular-dynamics method. *J. Appl. Phys.* 52, 7182–7190.

(60) Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh Ewald method. *J. Chem. Phys.* 103, 8577–8577.

(61) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10089.

(62) Hockney, R. W., and Eastwood, J. W. (1988) *Computer simulation using particles*, 1st ed., Adam Hilger, Bristol, U.K.

- (63) Miyamoto, S., and Kollman, P. A. (1992) Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962.
- (64) Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINC: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472.
- (65) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4, 435–447.
- (66) Berendsen, H. J. C., van der Spoel, D., and Vandrunen, R. (1995) GROMACS: A message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* 91, 43–56.
- (67) Klinger, M., Kuhn, M., Just, H., Stefan, E., Palmer, T., Freissmuth, M., and Nanoff, C. (2002) Removal of the carboxy terminus of the A_{2A}-adenosine receptor blunts constitutive activity: Differential effect on cAMP accumulation and MAP kinase stimulation. *Naunyn-Schmiedeberg's Arch. Pharmacol.* 366, 287–298.
- (68) Safhi, M. M., Rutherford, C., Ledent, C., Sands, W. A., and Palmer, T. M. (2010) Priming of signal transducer and activator of transcription proteins for cytokine-triggered polyubiquitylation and degradation by the A_{2A} adenosine receptor. *Mol. Pharmacol.* 77, 968–978.
- (69) Angel, T. E., Chance, M. R., and Palczewski, K. (2009) Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8555–8560.
- (70) Pardo, L., Deupi, X., Dölker, N., López-Rodríguez, M. L., and Campillo, M. (2007) The role of internal water molecules in the structure and function of the rhodopsin family of G protein-coupled receptors. *ChemBioChem* 8, 19–24.
- (71) Angel, T. E., Gupta, S., Jastrzebska, B., Palczewski, K., and Chance, M. R. (2009) Structural waters define a functional channel mediating activation of the GPCR, rhodopsin. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14367–14372.
- (72) Freites, J. A., Tobias, D. J., von Heijne, G., and White, S. H. (2005) Interface connections of a transmembrane voltage sensor. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15059–15064.
- (73) Zhou, W., Flanagan, C., Ballesteros, J., Konvicka, K., Davidson, J., Weinstein, H., Millar, R., and Sealson, S. (1994) A reciprocal mutation supports helix 2 and helix 7 proximity in the gonadotropin-releasing hormone receptor. *Mol. Pharmacol.* 45, 165–170.
- (74) Liapakis, G., Chan, W. C., Papadokostaki, M., and Javitch, J. A. (2004) Synergistic contributions of the functional groups of epinephrine to its affinity and efficacy at the β_2 adrenergic receptor. *Mol. Pharmacol.* 65, 1181–1190.
- (75) Vanni, S., Neri, M., Tavernelli, I., and Rothlisberger, U. (2010) A conserved protonation-induced switch can trigger "ionic-lock" formation in adrenergic receptors. *J. Mol. Biol.* 397, 1339–1349.
- (76) Selent, J., Sanz, F., Pastor, M., and De Fabritiis, G. (2010) Induced effects of sodium ions on dopaminergic G-protein coupled receptors. *PLoS Comput. Biol.* 6, e1000884.
- (77) Lenzi, O., Colotta, V., Catarzi, D., Varano, F., Poli, D., Filacchioni, G., Varani, K., Vincenzi, F., Borea, P. A., Paoletta, S., Morizzo, E., and Moro, S. (2009) 2-Phenylpyrazolo[4,3-d]pyrimidin-7-one as a new scaffold to obtain potent and selective human A₃ adenosine receptor antagonists: New insights into the receptor-antagonist recognition. *J. Med. Chem.* 52, 7640–7652.
- (78) Gao, Z. G., Jiang, Q., Jacobson, K. A., and Ijzerman, A. P. (2000) Site-directed mutagenesis studies of human A_{2A} adenosine receptors: Involvement of Glu(13) and His(278) in ligand binding and sodium modulation. *Biochem. Pharmacol.* 60, 661–668.
- (79) Ijzerman, A., van der Wenden, E. M., van Galen, P. J. M., and Jacobson, K. A. (1994) Molecular modelling of adenosine receptors. The ligand binding site on the rat adenosine A_{2A} receptor. *Eur. J. Pharmacol.* 268, 95–104.
- (80) Ijzerman, A. P., Kunzel, J. K. V. F. D., Kim, J., Qiaoling, J., and Jacobson, K. A. (1996) Site-directed mutagenesis of the human adenosine A_{2A} receptor. Critical involvement of Glu13 in agonist recognition. *Eur. J. Pharmacol.* 310, 269–272.
- (81) Morozov, A. V., Kortemme, T., Tsemekhman, K., and Baker, D. (2004) Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6946–6951.
- (82) Jiang, Q., Lee, B. X., Glashofer, M., van Rhee, A. M., and Jacobson, K. A. (1997) Mutagenesis reveals structure-activity parallels between human A_{2A} adenosine receptors and biogenic amine G protein-coupled receptors. *J. Med. Chem.* 40, 2588–2595.
- (83) Fredholm, B. B., Ijzerman, A. P., Jacobson, K. A., Klotz, K. N., and Linden, J. (2001) International Union of Pharmacology. XXV. Nomenclature and classification of adenosine receptors. *Pharmacol. Rev.* 53, 527–552.
- (84) Yaziji, V., Rodriguez, D., Gutierrez-de-Teran, H., Coelho, A., Caamano, O., Garcia-Mera, X., Brea, J., Loza, M. I., Cadavid, M. I., and Sotelo, E. (2011) Pyrimidine derivatives as potent and selective A₃ adenosine receptor antagonists. *J. Med. Chem.* 54, 457–471.
- (85) Jaakola, V. P., Lane, J. R., Lin, J. Y., Katritch, V., Ijzerman, A. P., and Stevens, R. C. (2010) Ligand binding and subtype selectivity of the human A_{2A} adenosine receptor: Identification and characterization of essential amino acid residues. *J. Biol. Chem.* 285, 13032–13044.
- (86) Xu, F., Wu, H., Katritch, V., Han, G. W., Jacobson, K. A., Gao, Z. G., Cherezov, V., and Stevens, R. C. (2011) Structure of an Agonist-Bound Human A_{2A} Adenosine Receptor. *Science*.
- (87) Kim, S. K., Gao, Z. G., Rompaey, P. V., Gross, A. S., Chen, A., Calenbergh, S. V., and Jacobson, K. A. (2003) Modeling the Adenosine Receptors: Comparison of the Binding Domains of A_{2A} Agonists and Antagonists. *J. Med. Chem.* 46, 4847–4859.
- (88) Kim, J., Wess, J., van Rhee, A. M., Schöneberg, T., and Jacobson, K. A. (1995) Site-directed Mutagenesis Identifies Residues Involved in Ligand Recognition in the Human A_{2A} Adenosine Receptor. *J. Biol. Chem.* 270, 13987–13997.
- (89) Floquet, N., M'Kadmi, C., Perahia, D., Gagne, D., Berge, G., Marie, J., Baneres, J. L., Galleyrand, J. C., Fehrentz, J. A., and Martinez, J. (2010) Activation of the ghrelin receptor is described by a privileged collective motion: A model for constitutive and agonist-induced activation of a sub-class A G-protein coupled receptor (GPCR). *J. Mol. Biol.* 395, 769–784.
- (90) Schwartz, T. W., Frimurer, T. M., Holst, B., Rosenkilde, M. M., and Eling, C. E. (2006) Molecular mechanism of 7TM receptor activation: A global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* 46, 481–519.
- (91) Rivkees, S. A., Barbhuiya, H., and Ijzerman, A. P. (1999) Identification of the Adenine Binding Site of the Human A₁ Adenosine Receptor. *J. Biol. Chem.* 274, 3617–3621.
- (92) Jiang, Q., Van Rhee, A., Kim, J., Yehle, S., Wess, J., and Jacobson, K. (1996) Hydrophilic side chains in the third and seventh transmembrane helical domains of human A_{2A} adenosine receptors are required for ligand recognition. *Mol. Pharmacol.* 50, 512–521.
- (93) Nygaard, R., Frimurer, T. M., Holst, B., Rosenkilde, M. M., and Schwartz, T. W. (2009) Ligand binding and micro-switches in 7TM receptor structures. *Trends Pharmacol. Sci.* 30, 249–259.

Paper V

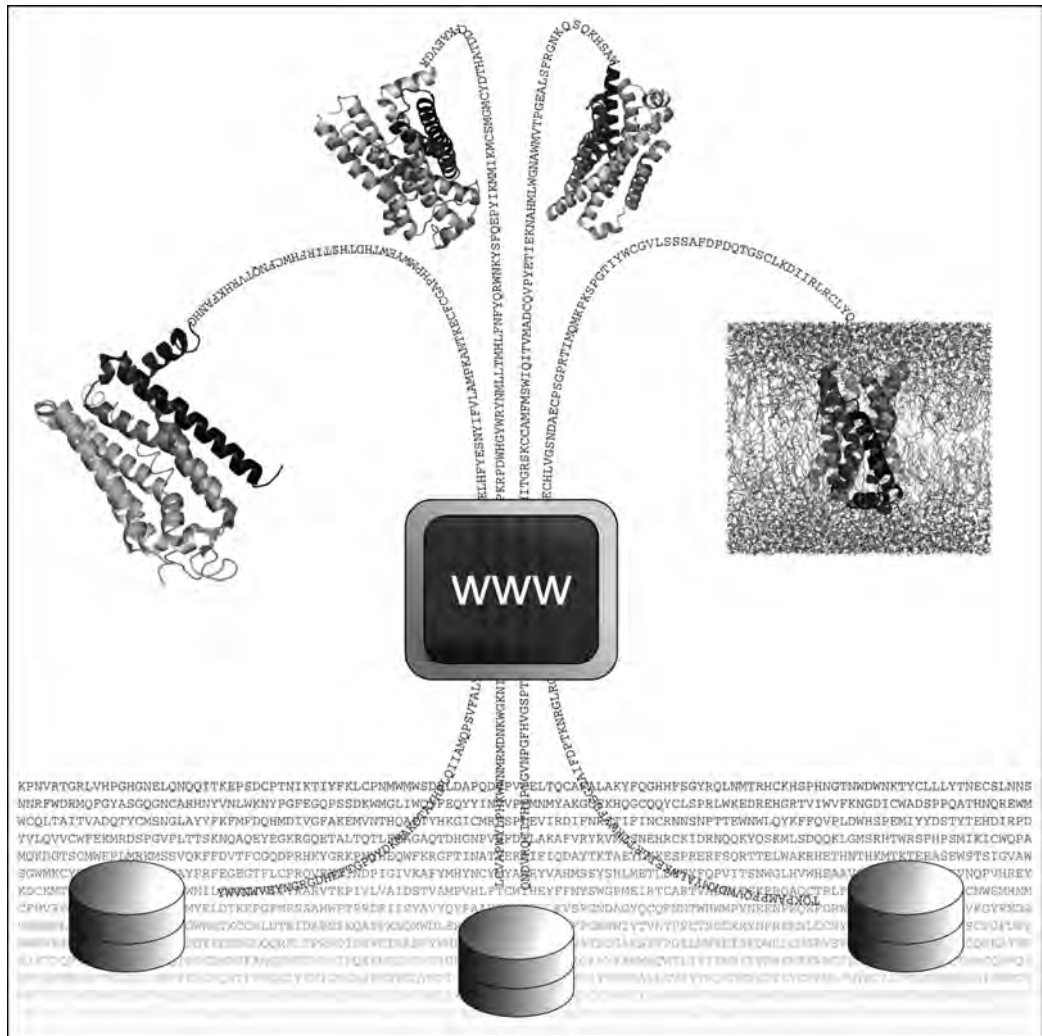
Rodríguez D., Bello X. and Gutiérrez-de-Terán H.

Molecular Modeling of G Protein-Coupled Receptors Through the Web

Molecular Informatics, **2012** 31:334-41

Molecular Modelling of G Protein-Coupled Receptors Through the Web

David Rodríguez,^[a] Xavier Bello,^[a] and Hugo Gutiérrez-de-Terán^{*[a]}



Abstract: With the recent crystallization of several G Protein-Coupled receptors (GPCRs), homology modelling and all atom molecular dynamics (MD) simulations have proven their usefulness for exploring the structure and function of this superfamily of membrane receptors. Subsequently, automated computational protocols have been implemented as web-based servers in the recent years to produce reliable models of GPCRs, providing partial or global solutions for the structural characterization and molecular simulation

Keywords: Database · G Protein-coupled receptor · Molecular dynamics · Structural bioinformatics · Web server

1 Introduction

G protein-coupled receptors (GPCRs) are transmembrane proteins in charge of signal transduction across cellular membranes. This superfamily of receptors represents approximately the 2% of the genes in the human genome,^[1] and constitute the main drug target (approximately 30% of the marketed drugs). The first crystal structure of a GPCR (bovine rhodopsin) was released in 2000,^[2] but it was not until 2007 that important advances in the receptor stabilization^[3] have allowed an extraordinary increase of experimental crystal structures of other receptors. As a result, representative members of six GPCR subfamilies, all belonging to class-A (rhodopsin-like) receptors,^[1b] are currently deposited in the PDB in their inactive conformation. In addition, members of three of these subfamilies (rhodopsin, adenosine and β -adrenergic receptors) have also been crystallized in active-like conformations. This structural information, combined with the conserved topology in the superfamily consisting on a seven-transmembrane helical (7TMH) bundle, facilitates the generation of accurate computer-derived molecular models of the remaining class-A GPCRs, ready to use in structure-based drug design projects.^[4] In this scenario, the scientific community is in need of automated methods to elucidate the 3D structure of the majority of the receptors of pharmacological interest. Moreover, there is an increasing demand of molecular dynamics (MD) simulations in this field, given its demonstrated utility in complementing the knowledge gained with experimental and modelled structures.^[5]

A number of computational procedures have been proposed for the modelling of GPCRs, as collected in the recent Critical Assessment of GPCR Structure Modelling and Docking (GPCR Dock) competitions.^[6] We here focus on those methods currently available as web servers as well as in databases of already generated GPCR models, since they offer an added value to researchers that, being interested in GPCR structure, have no prior experience in molecular modelling.

of GPCRs. These dedicated modelling services represent an attractive tool for the broader community of public researchers and pharmaceutical companies, in order to assist in the structure-based drug design of GPCRs. We here collect and analyze the existing web servers, among which a previously unreported service, GPCR-ModSim, offers for the first time full atom MD simulations in the pipeline for GPCR molecular modelling.

2 State-of-the-Art

2.1 Molecular Modelling of GPCRs

In the last years several molecular modelling methodologies have been adapted to obtain high quality models of GPCR structures, and published as software add-ons or web servers. The selection of the most appropriate protocol behind each service is a delicate issue, and is conditioned by several factors as it is summarized below:

- *Selection of the most appropriate modelling technique:* Looking at the recent GPCR Dock competitions,^[6] comparative modelling techniques emerge as the most popular methods, although in many cases limited by the availability of an appropriate template with sufficient homology (see next point). Motivated by this limitation, the so-called topology or first-principles modelling techniques have been developed.^[7] In essence, these methods strongly rely on computational biophysical methods to build the different TMH, pack them recognizing the conserved topology of a 7TMH bundle, and again use biophysical calculations [i.e., Monte Carlo (MC) or MD simulations] to further add loop regions and refine the obtained model. Some of these methods have proven their utility in the pharmaceutical design of GPCR ligands,^[8] but generally speaking their performance is comparable to the template-based methods.^[7b]
- *Selection of the most appropriate template:* this step is a complex issue that has been explored specifically for GPCRs.^[9] Although the number of crystal structures is expected to grow in the forthcoming years, the structural coverage of the phylogenetic tree of human GPCRs is still sparse.^[1,10] Frequently the overall sequence identity with any of the available templates is below the 30% threshold recommended for accurate modelling in membrane proteins.^[11]

[a] D. Rodríguez, X. Bello, H. Gutiérrez-de-Terán
Fundación Pública Galega de Medicina Xenómica (FPGMX),
Hospital Clínico Universitario de Santiago
pl-2, A Choupana s/n, 15706-Santiago de Compostela, Spain
phone/fax: +34-981951491/73
*e-mail: hugo.teran@usc.es

- *Model refinement*: given the conserved 7TMH topology of GPCRs, accurate modelling of the remaining extracellular (EL) and intracellular (IL) loops remains elusive. Thus, it is frequent to refine these regions using either *ab initio* modelling or dedicated databases for comparative loop modelling. Once a full model for the query GPCR is proposed, MD simulation considering a model of the lipid bilayer has become a state-of-the-art technique for model refinement,^[5] although this is often treated as a post-processing step after the modelling stage.
- *Automation of the process*: The availability of an automated method to model and simulate GPCRs is highly demanded. Otherwise, the general process (i.e., sequence alignments, model evaluation and selection, or the preparation of MD simulations) is tedious and must be assisted by a molecular modeller with additional background in GPCR biology.

David Rodríguez has a MSc in Biochemistry and obtained a MPhil in Molecular Medicine from the University of Santiago de Compostela in 2009, for the study of G Protein-Coupled Receptors (GPCRs) involved in neuropsychiatry disorders. He is currently a PhD Student at the FPGMX, focused in the molecular modelling, computer-aided ligand design and computational simulation of GPCRs. Other interests include chemoinformatics and virtual screening.



Xabier Bello has received his MSc in Biology at the University of Santiago de Compostela in 2008. He is currently pursuing his PhD at the same university, on the subject of population and evolutionary genomics of mobile elements. His interests are focused in the field of bioinformatics, and he is currently working as a bioinformatics assistant at the FPGMX.



Hugo Gutiérrez-de-Terán received his PhD in Health and Life Sciences (University Pompeu Fabra, 2004). After a postdoctoral period at Uppsala University, he joined the FPGMX as a Parga Pondal Research Fellow, where he has started the structural bioinformatics unit in 2008. He has focused on the application of molecular modelling and molecular simulation techniques in drug design projects, with particular emphasis on membrane receptors.



2.2 Databases of Computer-Derived Models of GPCRs

There are three dedicated databases of already generated three-dimensional molecular models of GPCRs (Table 1). This is quite convenient for researchers with no experience at all in molecular modelling, since the retrieved structures can be directly used, with some confidence, for several explorative tasks needed for the design or analysis of biochemical experiments. The GPCRDB^[12] is a comprehensive information system devoted to GPCRs, which is continuously updated since its first release in 1996. It includes a homology model of every receptor in the database, generated with WHATIF and YASARA. The model is offered together with the identification of the template used and the regions considered for homology modelling. The GPCR-SSFE database,^[13] published in 2010, includes 3D models of the 7TMH bundle for every GPCR annotated in the GPCRDB. For each sequence, a careful selection of multiple templates using a specific HMM algorithm^[9a] is used as a basis for the generation of homology models of the 7TMH bundle with Modeller.^[14] Alternative models are offered for each GPCR, depending on the particular selection of templates. Although not strictly required, ad hoc addition of the lacking connecting loops and further MD refinement is recommended by the authors.^[13] On the contrary, the database TASSER-907-GPCRs^[15] offers full-length 3D models for the subset of the 907 human GPCRs with less than 500 aminoacids, as deposited in the GPCRDB in 2006. A threading assembly refinement method, predecessor of I-TASSER (see next point),^[16] has been used in the model generation. Although this method allows reasonable modelling of loop regions under a given size,^[15] the presence of long loop segments in some of the deposited structures has led to mismodelled regions (i.e., long IL2 in aminergic receptors placed in the area of the lipid bilayer). The only version of this database dates from 2006, when only one crystal GPCR (bovin rhodopsin) was available. Very recently (October 2011), the GPCR-I-TASSER service has extended the generic threading modelling server I-TASSER^[16] by incorporating the protein-membrane interactions and the mutagenesis restraints compiled in the Restraint Database (GPCRRD)^[17] into its knowledge-based force field. With this strategy, the authors have created and stored structural models of all GPCRs in the human genome.

Finally, we should mention comprehensive databases of protein structures like ModBase,^[18] or SWISS-MODEL^[19] since they also include modelled structures for any sequenced GPCR. The deposited models have been obtained by homology modelling using Modeller^[14] (ModBase) or Swiss-model routines,^[19] but in the two cases the automated selection of templates was solely based on sequence homology, and the allowed decomposition of the sequence in order to build partial models importantly limits their applicability in the GPCR field.

Table 1. List of Databases and web services that can be used, to a different extent, for the obtention of three dimensional models of GPCRs.

Server name	URL	Alignment	Model	Loops	Dynamics	Comments
GPCRDB ^[a]	http://www.gpcr.org/7tm/	Y	Y	–	–	Database of GPCR models, based on multiple sequence alignments and generated with WHATIF and YASARA
GPCR-SSFE Database ^[a]	http://www.ssfa-7tmr.de/ssfe/index.php	Y	Y	N	N	Database of 3D models, and iterative model-builder of requested sequences. Different multiple template selections.
TASSER 907 GPCRs ^[a]	http://cssb.biology.gatech.edu/skolnick/files/gpcr/gpcr.html	–	–	–	–	Database of 3D-models of human GPCRs. Relatively outdated
GPCR-ITASSER ^[a]	http://zhanglab.ccmb.med.umich.edu/GPCR RD/GPCR ITASSER.html	–	Y	Y	N	Generation of GPCR models based on threading and spatial restraints from GPCR RD.
GPCR-ModSim ^[a]	http://gpcr.usc.es	Y	Y	Y	Y	All stages of GPCR modelling are included; allows user intervention: MD equilibration in POPC bilayer.
SwissModel ^[b]	http://swissmodel.expasy.org/	Y	Y	N	N	General purpose (not devoted to GPCR) database and iterative service.
Modbase ^[b]	https://modbase.compbio.ucsf.edu/scgi/mod-web.cgi	Y	Y	Y	N	General purpose (not devoted to GPCR) database and iterative service.
SuperLooper ^[c]	http://bioinf-applied.charite.de/superlooper/	N	N	Y	N	Loop modelling for transmembrane proteins. Detection of membranaral planes.
MEDELLER ^[c]	http://portal.stats.ox.ac.uk/userdata/proteins/medeller/gui/home.pl?app=MEDELLER	N	Y	Y	N	Performs comparative TM and loop modelling. User must select and upload template and query-template alignment.
CHARMM-GUI ^[c]	http://www.charmm-gui.org/	N	N	N	Y	Performs membrane insertion and initial setup for MD simulations with CHARMM.

[a] GPCR-dedicated web servers; [b] general purpose modelling servers; [c] additional web-based tools that complement some modelling/simulation steps.

2.3 Web Services Dedicated to the Computational Modelling of GPCRs

The GPCR-SSFE offers an associated, interactive web service to model any GPCR sequence upon demand, using the same modelling protocol followed in the creation of the database repository. Given the comprehensive coverage of this associated database, the interactive service is particularly useful to generate models of alternative sequences, i.e., receptors containing point mutations or artificial chimera constructs frequently used in molecular biology. The GPCR-ITASSER interactive server must be used to generate models of any non-human GPCR, which are not included in its associated database, in addition to the other potential applications (i.e., the investigation of alternate sequences as depicted below). It is worth noting that, despite methodological differences (i.e., *ab initio* vs. comparative modelling) this procedure is more computationally demanding than the GPCR-SSFE, and indeed the generation of a single model can take several days. As it occurs with the databases of structural models, the general purpose modelling

services such as ModWeb^[18] or Swiss-Model^[19] can generate reasonable models in easy cases, where a high homologous GPCR template exists, but are of limited applicability in more complex situations unless guided by expert human intervention.

Attending to the modelling protocols depicted in section 2.1, the web services described above cover to a different extent the first stages of the GPCR modelling (see Figure 1), that is: *i*) the selection of the best template(s), *ii*) the generation of a template-query alignment, *iii*) the initial modelling of either the 7TMH region (GPCR-SSFE) or even including *iv*) a loop refinement or the EL and IL regions (GPCR-ITASSER). Alternatively, one can also model a full GPCR using a combination of web servers that specifically perform some of these steps. One possibility is to use the MEDELLER server^[20] which implements a homology-based coordinate generation method optimized for membrane proteins, provided that an accurate query-template alignment has been previously generated by the user (i.e., as obtained from the GPCRDB^[12]). The predicted membrane insertion of the template protein is used to define the core

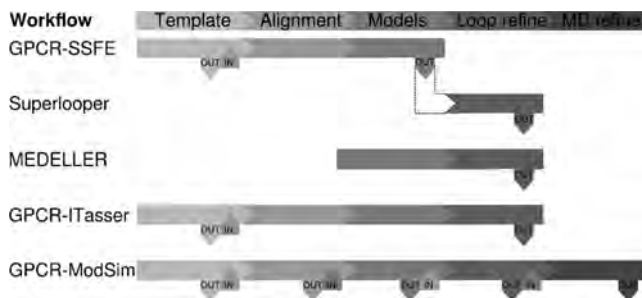


Figure 1. General workflow to obtain a three dimensional model of a GPCR from its sequence. The stages covered by the each of the web servers discussed in the text are indicated. The communication with the user is indicated with arrows.

region which is modelled by homology, while the remaining loop regions are built on the basis of a loop modelling algorithm. Further refinement of the generated models is advised in order to correct local structural differences.^[20] Also interesting is the SuperLooper web server for loop modelling.^[21] This service needs as input a structural model of the transmembrane bundle the receptor. Thereafter, it detects the membrane delimiting regions of the provided model, to build each of the loop regions as specified by the user, by means of an internal loop database.^[21] Thus, a combination of GPCR-SSFE + SuperLooper can be used as a 2-stepwise protocol to obtain full models of a GPCR^[13] (see Figure 1).

Despite the modelling protocol followed, a model assessment and refinement is advised. ProQM^[22] is a server for model quality assessment adapted for TM proteins, with proved utility in GPCRs. It outperforms other consensus-based scoring functions, thus being an important aid to detect regions that need further optimization. Finally, if MD refinement is attempted, the initial steps for setting up the protein-membrane system can be simplified with the CHARMM-GUI web service.^[23]

2.4 From Sequence to MD Refinement: GPCR-ModSim

Recently, a web-based server that performs for the first time the whole pipeline of GPCR modelling, including refinement with MD simulations was released: GPCR-ModSim (Table 1) offers a simple web interface that performs all the stages of GPCR modelling (see Figure 1) with minimum user intervention. Structural alignments for inactive- and active-like GPCR crystal structures were obtained with the CAlign algorithm. The derived multiple sequence alignments (MSA) were manually checked and stored as “inactive” and “active” template profiles. After uploading the query sequence, the user selects the desired profile (i.e. “inactive” or “active”) and a MSA of templates-query sequences is generated, which can be inspected and manually edited through a Jalview applet. The individual sequence identity percentages of the query GPCR with each template

are offered for both the whole sequence and the 7TMH bundle. Moreover, sequence identity on each of the 15 regions defined in the GPCR conserved topology (i.e., 7TMHs, 3ELs, 3ILs, plus amino and carboxy termini) is separately examined for each query-template combination. This information and/or additional knowledge of the system (i.e., SAR or site-directed mutagenesis studies), can be used to modify the server’s default selection of the best template for the homology modelling phase (which is done on the basis of the highest 7TMH sequence identity). Thereafter, up to 10 different models are generated using Modeller 9v8,^[14] based on the selected template and allowing the user to restrain additional disulfide bridges (apart from the conserved one between TM3 and EL2 in class-A GPCRs). The models are ranked by DOPE-HR scoring, and the model with the lowest (most favoured) value is proposed for further refinement. However, the user can change this selection after visual inspection or considering the stereochemical evaluation reports provided with each model. The best initial model is then (optionally) submitted to a loop refinement process, using the LoopModel routine as implemented in Modeller.^[24,14b] The user defines the precise loop regions to be remodelled, and indicates (if any) optional additional disulfide bridges. A Jmol applet aids in the visual inspection of the models in both modelling steps described above. The loop-refined models can be further assessed using the same indicators than with the initial models (i.e., DOPE-HR scoring function, stereochemical quality, visual inspection) and a final structure is selected for MD refinement. The modelled GPCR is then inserted in an atomistic model of the cellular membrane, and the system is equilibrated with all-atom MD simulations using GROMACS.^[25] The MD phase is fully automated as adapted from ref. [26] and consists in *i)* the insertion of the receptor in an hexagonal prism-shaped simulation box, with TM helices parallel to the *z* axis, *ii)* solvation with a pre-equilibrated hydrated POPC bilayer and *iii)* neutralization of the system with explicit counter ions, as estimated at physiological pH. A MD equilibration protocol follows for a total simulation time of 5 ns, under periodic boundary conditions (see Ref. [26] for

details). Finally, the main output files are provided together with several reports and all the necessary files and instructions to extend the MD simulation. The GPCR-ModSim server stores and organizes the generated data for each project (one per query sequence), that the user can easily browse and download for 7 days.

3 Recent Applications and Limitations

Despite the increased popularity of biological databases, we have found little examples in the literature about the

utility of the specialized databases of GPCR models. This is understandable in the most recent databases, i.e., GPCR-SSFE,^[13] GPCR-ITASSER^[17] and even the models as recently deposited in the GPCRDB.^[12] For the most life-long database, the TASSER-907-GPCRs released in 2006,^[15] some interesting use has been documented: Wang and Duan retrieved the deposited model of CCR5, and further refined it with MD simulations, before exploring the binding mode of small molecule inhibitors.^[27] Alternatively, the deposited structures can be taken as a gold standard to compare newly generated models, as it has happened with different sphingosine 1-phosphate (S1P) receptor family.^[28]



Figure 2. Molecular models of the h1 receptor, as generated with each dedicated web server using default settings. Each model is aligned to the experimental structure (green, PDB code 3rze) with the corresponding *RMSD* values and colour codes indicated in the accompanying table.

In the two editions of the GPCR Dock competitions, celebrated in 2008 and 2010,^[6] the only available dedicated web-server at that time, I-TASSER,^[16] was used without further modifications (see the contributing model of Horst et al. in reference [6a] and KIAS group in reference [6b]), or even its modelling protocol was adapted in what recently became the GPCR-ITASSER server (see contribution from UMich-Zhang group in reference [5b]). Indeed, the Zhang group has reported one of the best predictions for the CXCR4 structure in complex with the cyclic peptide,^[6a] demonstrating the utility of this ab initio algorithm in the prediction of disordered loop regions. In the GPCR Dock 2008 competition, predictions of A_{2A}AR structure by Rodríguez and Gutiérrez-de-Terán were ranked within the top 10% of all models (codes mod3oak and mod7xmu), by employing the basic elements of the pipeline that would be lately used in the initial model building of the GPCR-ModSim server. This server was recently employed to select the best template for the modelling of the human neuropeptide receptor Y2.^[29]

Besides the proven complete automation of the process, the three dedicated web servers are somehow flexible, allowing the user to input previous knowledge on the system (see Figure 1). In the case of GPCR-SSFE, the flexibility is limited to the generation of several models for each query sequence, depending on the selection of the templates for the different TMHs. The GPCR-ITASSER server^[17] automatically selects the template and the experimental distance restraints, but the user is referred to the more generic I-TASSER associated server^[16] if a fine-tuning of these parameters is desired. The GPCR-ModSim is the most flexible server, allowing the user to decide on every step involved in the modelling pipeline: selection of the template, manual edition of the query-template alignment, selection of the most appropriate 3D-model, and definition of the loop regions to be refined. The insertion in an atomistic model of the cellular membrane and further equilibration with MD simulations is an exclusive feature of this server, which can also be used as a stand-alone service by uploading a GPCR structure obtained by external methods.

4 Outlook

The generation of accurate molecular models of GPCRs is actually a limiting step to proceed with a deeper structural and conformational analysis of the majority of human receptors, including structure-based drug design projects.^[4,10] In this sense, the databases of already generated GPCR models have a potential interest to users with no experience in macromolecular modelling. The structures deposited in these databases allow simple in silico analysis that can complement wet lab experiments, including: *i)* the three dimensional location of specific residues, in order to infer potential roles in protein architecture or even to anticipate functional effects of mutations, *ii)* the location of the

binding site crevice, or *iii)* the structural alignment of all members of a GPCR family. However, it is often recommended to refine, or at least check carefully these crude models before proceeding to more complex studies,^[13] such as virtual screening or structure-based drug design.

A number of web sites exist, offering interesting services that automate or reduce, to a different extent, the stages needed to generate a computational model of a GPCR (Table 1). Three services, namely GPCR-SSFE, GPCR-ITASSER and GPCR-ModSim, emerge as the only GPCR dedicated web servers. The three servers can generate accurate 3D-models of the recently crystallized hH1 histamine receptor, as shown in Figure 2, with minimum user intervention.

The degree of flexibility of each server is different, as illustrated in Figure 1, and it is worth indicating that some added human intervention, in particular if additional experimental information is provided, can greatly improve the quality of the models.^[4] The most flexible server, GPCR-ModSim, is also the only service that offers a stand-alone protocol for all modelling stages including full-atom MD equilibration. This is an added value since MD simulations are becoming a very common technique to further refine experimental and modelled structures,^[5] or even enrich virtual screenings.^[30] Finally, the expected increase in the number of GPCR templates^[10] will not only be a benefit itself, but will enhance the quality of computer-derived models in the forthcoming years.

Acknowledgements

This research is funded by *Xunta de Galicia, Consellería de Sanidade* (PS09/63) and the *Spanish Ministry of Science and Innovation* (SAF2011-30104). HGT is a *Parga Pondal* research fellow (*Xunta de Galicia*) and DR is recipient of a predoctoral grant from the Fondo de Investigación Sanitaria (ISCIII).

References

- [1] a) R. Fredriksson, M. C. Lagerström, L.-G. Lundin, H. B. Schiöth, *Mol. Pharmacol.* **2003**, *63*, 1256–1272; b) The A–F classification system for GPCRs is adopted, where the superfamily of GPCRs is divided into 5 classes (A to F), each of which is composed by several subfamilies.
- [2] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, M. Miyano, *Science* **2000**, *289*, 739–745.
- [3] D. M. Rosenbaum, V. Cherezov, M. A. Hanson, S. G. F. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, X.-J. Yao, W. I. Weis, R. C. Stevens, B. K. Kobilka, *Science* **2007**, *318*, 1266–1273.
- [4] M. Congreve, C. J. Langmead, J. S. Mason, F. H. Marshall, *J. Med. Chem.* **2011**, *54*, 4283–4311.
- [5] a) D. M. Rosenbaum, C. Zhang, J. A. Lyons, R. Holl, D. Aragao, D. H. Arlow, S. G. Rasmussen, H. J. Choi, B. T. Devree, R. K. Sunahara, P. S. Chae, S. H. Gellman, R. O. Dror, D. E. Shaw, W. I. Weis, M. Caffrey, P. Gmeiner, B. K. Kobilka, *Nature* **2011**, *469*, 236–240; b) R. O. Dror, D. H. Arlow, D. W. Borhani, M. Ø.

- Jensen, S. Piana, D. E. Shaw, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4689–4694.
- [6] a) M. Michino, E. Abola, C. L. Brooks, J. S. Dixon, J. Moul, R. C. Stevens, *Nat. Rev. Drug Discov.* **2009**, *8*, 455–463; b) I. Kufareva, M. Rueda, V. Katritch, R. C. Stevens, R. Abagyan, *Structure* **2011**, *19*, 1108–1126.
- [7] a) N. Vaidehi, W. B. Floriano, R. Trabanino, S. E. Hall, P. Freddolino, E. J. Choi, G. Zamanakos, W. A. Goddard, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12622–12627; b) M. Michino, J. Chen, R. C. Stevens, C. L. Brooks, *Proteins* **2010**, *78*, 2189–2201.
- [8] O. M. Becker, D. S. Dhanoa, Y. Marantz, D. Chen, S. Shacham, S. Cheruku, A. Heifetz, P. Mohanty, M. Fichman, A. Sharadendu, R. Nudelman, M. Kauffman, S. Noiman, *J. Med. Chem.* **2006**, *49*, 3116–3135.
- [9] a) C. L. Worth, G. Kleinau, G. Krause, *PLoS One* **2009**, *4*, e7011; b) J. C. Mobarec, R. Sanchez, M. Filizola, *J. Med. Chem.* **2009**, *52*, 5207–5216.
- [10] V. Katritch, V. Cherezov, R. C. Stevens, *Trends Pharmacol. Sci.* **2012**, *33*, 17–27.
- [11] L. R. Forrest, C. L. Tang, B. Honig, *Biophys. J.* **2006**, *91*, 508–517.
- [12] B. Vroiling, M. Sanders, C. Baakman, A. Borrmann, S. Verhoeven, J. Klomp, L. Oliveira, J. de Vlieg, G. Vriend, *Nucleic Acids Res.* **2011**, *39*, D309–319.
- [13] C. L. Worth, A. Kreuchwig, G. Kleinau, G. Krause, *BMC Bioinformatics* **2011**, *12*, 185.
- [14] a) A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779–815; b) The user must provide a valid Modeller license key, which can be obtained at <http://salilab.org/modeller/registration.shtml>.
- [15] Y. Zhang, M. E. Devries, J. Skolnick, *PLoS Comput. Biol.* **2006**, *2*, e13.
- [16] A. Roy, A. Kucukural, Y. Zhang, *Nat. Protoc.* **2010**, *5*, 725–738.
- [17] J. Zhang, Y. Zhang, *Bioinformatics* **2010**, *26*, 3004–3005.
- [18] U. Pieper, B. M. Webb, D. T. Barkan, D. Schneidman-Duhovny, A. Schlessinger, H. Braberg, Z. Yang, E. C. Meng, E. F. Pettersen, C. C. Huang, R. S. Datta, P. Sampathkumar, M. S. Madhusudhan, K. Sjolander, T. E. Ferrin, S. K. Burley, A. Sali, *Nucleic Acids Res.* **2011**, *39*, D465–474.
- [19] F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, T. Schwede, *Nucleic Acids Res.* **2009**, *37*, D387–392.
- [20] S. Kelm, J. Shi, C. M. Deane, *Bioinformatics* **2010**, *26*, 2833–2840.
- [21] P. W. Hildebrand, A. Goede, R. A. Bauer, B. Gruening, J. Ismer, E. Michalsky, R. Preissner, *Nucleic Acids Res.* **2009**, *37*, W571–574.
- [22] A. Ray, E. Lindahl, B. Wallner, *Bioinformatics*, **2010**, *26*, 3067–3074.
- [23] S. Jo, T. Kim, W. Im, *PLoS One* **2007**, *2*, e880.
- [24] A. Fiser, R. K. Do, A. Sali, *Protein Sci.* **2000**, *9*, 1753–1773.
- [25] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- [26] D. Rodríguez, A. Piñero, H. Gutiérrez-de-Terán, *Biochemistry* **2011**, *50*, 4194–4208.
- [27] T. Wang, Y. Duan, *J. Mol. Graph. Model.* **2008**, *26*, 1287–1295.
- [28] a) S. C. Schurer, S. J. Brown, P. J. Gonzalez-Cabrera, M. T. Schaeffer, J. Chapman, E. Jo, P. Chase, T. Spicer, P. Hodder, H. Rosen, *ACS Chem. Biol.* **2008**, *3*, 486–498; b) T. C. Pham, J. I. Fells, Sr., D. A. Osborne, E. J. North, M. M. Naor, A. L. Parrill, *J. Mol. Graph. Model.* **2008**, *26*, 1189–1201.
- [29] H. Fallmar, H. Kerberg, H. Gutiérrez-de-Terán, I. Lundell, N. Mohell, D. Larhammar, *Neuropeptides* **2011**, *45*, 293–300.
- [30] M. Wada, E. Kanamori, H. Nakamura, Y. Fukunishi, *J. Chem. Inf. Model.* **2011**, *51*, 2398–2407.

Received: December 8, 2011

Accepted: January 6, 2012

Published online: February 8, 2012

Paper VI

Rodríguez D. and Gutiérrez-de-Terán H.

Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor

Proteins, 2012 80(8):1919-28

SHORT COMMUNICATION

Characterization of the homodimerization interface and functional hotspots of the CXCR4 chemokine receptor

David Rodríguez and Hugo Gutiérrez-de-Terán*

Fundación Pública Galega de Medicina Xenómica, Hospital Clínico Universitario de Santiago, Santiago de Compostela E-15706, Spain

ABSTRACT

The recent crystallographic structures of the human chemokine CXC Receptor 4 (CXCR4) provide experimental evidence of a human G Protein-Coupled Receptor (GPCR) dimer in atomic detail. The CXCR4 homodimers reveal an unexpected dimerization mode involving transmembrane helices TM5 and TM6, which is examined here using all-atom molecular dynamics (MD) simulations in the physiological environment of a lipid bilayer. The bacteriophage T4 lysozyme (T4L), which was fused to the crystallized protein but absent in our simulations, is found to slightly affect the observed relative position of the protomers in the two dimers studied here, and consequently some rearrangements of the dimerization interface are proposed. In addition, the simulations provide further evidence about the role of the two stabilizing single point mutations introduced to crystallize the receptor. Finally, this work analyzes the structural and dynamic role of key residues involved both in ligand binding and in the infection process of HIV. In particular, the different side chain conformations of His113^{3,39} are found to influence the dynamics of the surrounding functional hotspot region being evaluated both in the presence and in the absence of the co-crystallized ligand IT1t. The analysis reported here adds valuable knowledge for future structure-based drug design (SBDD) efforts on this pharmacological target.

Proteins 2012; 80:1919–1928.
© 2012 Wiley Periodicals, Inc.

Key words: binding site; chemokine; CXCR4; GPCR; homodimer; HIV; hotspot residues; MD simulations.

INTRODUCTION

Signal transduction across cellular membranes is mainly controlled by the G protein-coupled receptors (GPCRs). In this family of seven transmembrane helical proteins, the binding of the extracellular physiological ligand stabilizes the active conformation of the receptor, leading to the cellular response.¹ Chemokine receptors are GPCRs which mediate the response of the cell to the extracellular levels of chemokines. Upon activation, these receptors regulate pathways that induce changes in cytoskeleton proteins, which are ultimately translated into the biological phenomenon of chemotaxis. The 20 known human chemokine receptors have been implicated in many pathological processes related to cell migration,

including inflammation and autoimmune diseases.² CXC Receptor 4 (CXCR4) is specifically activated by the chemokine CXCL12 (SDF-1),² and it participates in the membrane fusion and cell entry processes of HIV-1 in

Additional Supporting Information may be found in the online version of the article.

Grant sponsor: Xunta de Galicia, Consellería de Sanidade; Grant number: PS09/63; Grant sponsor: Ministerio de Ciencia e Innovación; Grant numbers: SAF2011-30104 and ICTS-2009-40.

*Correspondence to: Hugo Gutiérrez de Terán, Fundación Pública Galega de Medicina Xenómica, Complejo Hospitalario Universitario de Santiago (CHUS), Planta-2, A Choupana, s/n. E-15706, Santiago de Compostela, Spain.

E-mail: hugo.teran@usc.es

Received 2 February 2012; Revised 3 April 2012; Accepted 13 April 2012

Published online 19 April 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24099

the host cell,^{3–5} making it a promising target for the treatment of AIDS. In addition, this receptor has been associated with up to 23 different types of cancer.⁶

Following the efforts in GPCR crystallization of the last decade,⁷ two high-resolution crystallographic structures of inactive conformations of CXCR4, in complex with the antagonists IT1t (PDB code 3ODU) and the cyclic peptide CVX15 (PDB code 3OE0), have become available.⁸ As a major breakthrough, each crystal structure is a (CXCR4)₂ homodimer, where the orientation of the two protomers of CXCR4 shows a novel dimerization interface for a human GPCR, with potential biological relevance on the system.⁹ The crystal structure with the small-molecule inhibitor bound (2.5 Å resolution) has two protomers in the crystallographic asymmetric unit, whereas the structure in complex with the cyclic peptide (2.9 Å resolution) is a biological assembly as defined by the authors.⁸ The latter structure presents additional contacts between protomers in the intracellular side, as compared to the former. In both cases, the dimerization interface of both (CXCR4)₂ systems is located within transmembrane (TM) helices TM5 and TM6, in contrast with previous atomic-force microscopy (AFM) models of the rhodopsin (Rho) receptor. The interface of the corresponding (Rho)₂ homodimers is formed by TM4 and TM5,^{10, 11} and had been considered a standard for the GPCR superfamily of receptors.¹² However, TM6 also has potential as an interacting region for GPCR dimerization, according to bioinformatics predictions¹³ or bioluminescence resonance energy transfer (BRET) studies of CXCR4 dimers.¹⁴ Both homo- and hetero-dimerization of GPCRs has been shown to be of major importance in the biological function of the whole superfamily,^{15, 16} including the suggested physiological role for (CXCR4)₂ homodimers.¹⁷

Ligand-independent CXCR4 dimerization has been experimentally observed and characterized by fluorescence resonance energy transfer (FRET) and BRET techniques,^{18, 19} and is one of the phenomena that we will examine in the context of the new crystal structures of CXCR4 reported by Wu et al.⁸ Additional areas that deserve further exploration related to these crystal structures include the possible influence of the T4 bacteriophage lysozyme (T4L) fused to the intracellular loop (IL) 3 region of the receptors in the dimerization interface, and the effect of the physiological environment of a lipid bilayer on the receptor, which contrasts with the conditions used in the crystallization process. It is important to address these questions to hypothesize a biological significance for the experimentally observed dimerization interface, and extrapolate these findings toward the design of dual ligands with high therapeutic relevance.²⁰ Here, we approach these problems through a computational examination of the dimerization interface of the CXCR4 homodimer by adapting our recently reported protocol for the molecular dynamics (MD) simulations

of GPCRs²¹ to the case of receptor dimers. Additionally, we analyze the dynamics of some residues previously implicated by mutagenesis studies as relevant for the infection of HIV. In particular, we focused on a cluster of residues at TM4 and extracellular loop (EL) 2 that surround His113^{3,29} in TM3. The influence of the side chain conformation of His113^{3,29} in dynamic events is examined, with possible relevance in the SBDD of low molecular weight compounds, such as the co-crystallized antagonist IT1t.²²

MATERIALS AND METHODS

Modeling of missing regions and reversal of stabilizing mutations

The crystal structures of CXCR4 were retrieved from the Protein Data Bank (PDB codes 3ODU and 3OE0) and further refined before MD simulations. In addition, a wild-type (WT) version of each receptor was built by restoring the original side chain in each of the two stabilizing mutations present in the corresponding crystal structures by means of the Mutate tool of Maestro,²³ L125W^{3,41} in 3ODU and 3OE0, and T250P^{6,36} in 3OE0. In each system, bound ligands and lipid molecules were deleted. The T4L was removed from the IL3, missing residues and/or chain breaks at ILs were rebuilt with Modeller,²⁴ ranked according to DOPE-HR scoring,²⁵ and further refined and evaluated with Molprobit²⁶ and the PDB2PQR server.²⁷ Partial energy minimization followed with MacroModel.²³ Selected crystallographic water molecules were included before the proton addition and energy minimizations of the modeled regions (for further details see Supplementary Methods in the Supporting Information).

Molecular dynamics simulations

Membrane insertion and all MD simulations were performed with the GROMACS software.²⁸ A Python program was written to automate all the processes, an adaptation of the protocol for MD simulations of GPCRs previously developed in our group²¹ to the particular case of GPCR dimers. Briefly, the process consists of three steps: (i) each homodimer was introduced into the centre of an hexagonal prism-shaped simulation box, in a way that the TM helices were parallel to the z axis, every side box walls were located at a minimum distance of 35 Å and every top and bottom box wall was at a minimum distance of 12 Å for any protein atom (see Fig. S1 in the Supporting Information); (ii) a hydrated palmitoylcholine (POPC) gel-phase bilayer, pre-equilibrated at 260 K, was adjusted to the dimensions of the simulation box by removal of excess lipid and water molecules, including those overlapping with protein atoms; (iii) Cl⁻ ions were introduced to neutralize the net positive charge of the receptor, as estimated at physiological

Table I
Geometrical Descriptors Obtained in the MD Simulations of the Different (CXCR4)₂ Systems

Name	PDB code	Mutations	RMSD (Å)		Number of distances ^a		% Increase of BSA ^b
			Whole dimer ^c	Protomer B ^{c,d}	Close	Open	
(CXCR4) ₂ -Xray-mut	3ODU	L125W ^{2,41}	2.92 (± 0.13)	6.42 (± 0.49)	179	7	74.99%
(CXCR4) ₂ -Xray-wt	3ODU	—	2.96 (± 0.18)	5.35 (± 0.45)	125	37	30.87%
(CXCR4) ₂ -assembly-mut	3OE0	L125W ^{2,41} T250P ^{6,36}	2.20 (± 0.12)	2.90 (± 0.35)	45	28	12.12%
(CXCR4) ₂ -assembly-wt	3OE0	—	3.02 (± 0.14)	4.54 (± 0.33)	77	33	10.52%
(CXCR4-IT1t) ₂ -Xray-mut ^e	3ODU	L125W ^{2,41}	2.68 (± 0.16)	4.86 (± 0.41)	193	17	54.87%

^aNumber of closing and opening distances between residues of different protomers. Only residue pairs with distance variations >2 Å along the MD simulations are considered.

^bThe percentage of the increase of BSA between protomers is calculated as: $\text{fBSA} = \frac{\text{BSA}_{\text{end},5\text{ns}} - \text{BSA}_{\text{ini},5\text{ns}}}{\text{BSA}_{\text{ini},5\text{ns}}} \times 100$, where $\text{BSA}_{\text{end},5\text{ns}}$ and $\text{BSA}_{\text{ini},5\text{ns}}$ correspond to the average BSA of the final and initial 5 ns of the production phase, respectively.

^cAverage backbone RMSD (and associated standard error) calculated on the second half of the production phase of each simulation.

^dAll the snapshots are aligned considering the backbone of the other protomer (A), thus illustrating the relative movements between the two protomers along MD simulations.

^eSimulation in complex with the crystallographic ligand IT1t, with a production phase of 50 ns (the rest of simulations account for 100 ns).

pH. The final systems consist of ~100,000 atoms, from which ~73% belong to solvent molecules, ~18% to lipids, and ~8% to protein atoms. MD simulations were performed under the OPLSAA force field,²⁹ using the SPC water model³⁰ and applying the half- ϵ double-pairlist method³¹ to ensure compatibility of the lipid parameters adapted from Berger.³² The periodic boundary conditions were implemented with hexagonal prism-shaped boxes in the NPT ensemble. The simulations were carried out at a temperature of 310 K using a Nose-Hoover thermostat,³³ with independent coupling of the lipid molecules, the protein [together with the ligand in (CXCR4-IT1t)₂-Xray simulation], and the water–ion groups. A cutoff of 12 Å was employed for non-bonded interactions, applying the particle mesh Ewald (PME) method³⁴ for electrostatic interactions beyond such cutoff. More details on the MD parameters adopted for the simulation of GPCRs are provided in Reference 21. The equilibration protocol accounts for a total time of 10 ns, and was followed by 100 ns unrestrained MD simulations which were analyzed with several GROMACS utilities (see the Supplementary Methods in the Supporting Information).

The MD simulations of the (CXCR4-IT1t)₂-Xray-mut complex were conducted as described above, with the following additional parameters and options: one ligand molecule per protomer was kept, thus maintaining the 2:2 stoichiometry observed in the 3ODU crystal structure. Two setups were considered to account for the most probable protonation states of the ligand, as predicted by pK_a calculations³⁵; with a net charge of either +1 or +2, respectively, and varying the protonation state of the interacting residue Glu288^{7,39} accordingly (see Fig. S2 for details). OPLSAA force field parameters of the ligand IT1t were obtained with Macromodel²³ and translated to the GROMACS syntax using an *ad hoc* script. Such automated generation of ligand parameters has been successfully used for recent free energy calcula-

tions.³⁶ During the equilibration phase, the ligand was restrained as a side chain of the receptor (see Supplementary Methods of the Supporting Information). Unrestrained MD simulations followed for 50 ns in this system.

Molecular superimpositions, trajectory visualizations, and molecular images were performed with PyMOL v1.2r3 (<http://www.pymol.org>). The standard amino acid sequence number for the human CXCR4 receptor is indicated in normal text, with the Ballesteros and Weinstein residue numbering for GPCRs³⁷ provided as a superscript.

RESULTS AND DISCUSSION

Dynamics of the homodimerization interface

We considered four different versions of the (CXCR4)₂ system, accounting for the two proposed dimerization models and examined not only the crystallized receptors, which contain stabilizing mutations, but also the corresponding WT variants. The four systems will be depicted as (CXCR4)₂-Xray-mut (PDB ID 3ODU), (CXCR4)₂-assembly-mut (PDB ID 3OE0), (CXCR4)₂-Xray-wt, and (CXCR4)₂-assembly-wt as summarized in Table I. The dimerization interfaces of the parent (CXCR4)₂ crystal structures, located between the TM5 and TM6 regions, differ mainly in the intracellular half of the receptors, where the 3ODU (Xray) model presents a smaller contact surface area than the 3OE0 (assembly) model.⁸

The geometric analysis of the MD trajectories reveals that remarkable rearrangements in the dimerization interface occur in the first part of the simulations (10–40 ns), accounting for the relative approach between protomers, which is followed in all cases by a stable phase that lasts a minimum of 60 ns, suggesting that the corresponding dimerization interfaces have reached a signifi-

cant stationary state (see Fig. S3 in the Supporting Information). This initial rearrangement occurs to a lower extent in the (CXCR4)₂-assembly model, having the lowest average root mean squared deviation (RMSD) along the simulation time (Table I and Supplementary Fig. S3), consistent with the initial larger contact surface area of the parent crystal structure. Accordingly, the buried surface area (BSA) of the dimerization interface increases in all cases from only 10%, in the case of (CXCR4)₂-assembly-wt to >70% for (CXCR4)₂-Xray-mut (see Table I). The systems with the stabilizing mutations reach the stable phase faster than the corresponding WT version (Fig. S3), which indeed provides structural evidence of the effect of these mutations in the stabilization of the homodimer. The root mean square fluctuation (RMSF) analysis (Fig. 1) reveals remarkable fluctuations in the TM helices in the WT forms, in particular located in TM6 and TM7 of (CXCR4)₂-Xray-wt, (residues: 240–290), and TM1 (residues: 50–75) of (CXCR4)₂-assembly-wt, again suggesting a reduced stability for the WT receptor forms.

The location and magnitude of the most important differences in the dimerization interface has been identified with the aid of the contact map matrices shown in Figure 2(A). The (CXCR4)₂-Xray structures show the higher number of closing distances, consistent with the greater evolution of their BSA (see Table I) and the lower number of initial contacts of this dimerization model. In all cases, the relative approach of the protomers is mainly located in the intracellular side of the dimerization interface [Fig. 2(B)].

Measurement of the relative mobility of protomer B after an alignment of all MD snapshots to protomer A (see Table I and Fig. 3) provides evidence of additional differences in the behavior of the two crystallographic models of (CXCR4)₂. In the Xray (3ODU) structure, the TM4 from each of the two monomers come close to each other. This movement involves a relative approach of the TM5 of one protomer to the TM4 of the other (see Fig. 2), which is achieved by a relative rotation between the two protomers (Fig. 3). Remarkably, this event is not observed in the (CXCR4)₂-assembly simulations, where the relative orientation between TM4 and TM5 remains stable. Conversely, in the last system, the π -stacking interactions involving Tyr135^{3,51} (belonging to the conserved DRY motif) and His140^{3,56} (at IL2), already present in the corresponding 3OE0 crystal structure, drive a relative approach of the intracellular ends of TM3 and TM5. In addition, a set of novel but stable hydrogen bonds between protomers is identified with the MD simulations, which further stabilizes the intracellular contacts of the (CXCR4)₂-assembly structures. These are located in the region between the corresponding TM3 helices (Tyr135^{3,51} and His140^{3,56(IL2)}) and between TM4 and TM5 (residues Arg146^{4,35}, Ser225^{5,63} and Ser228^{5,66}), as depicted in Table II and Figure 4.

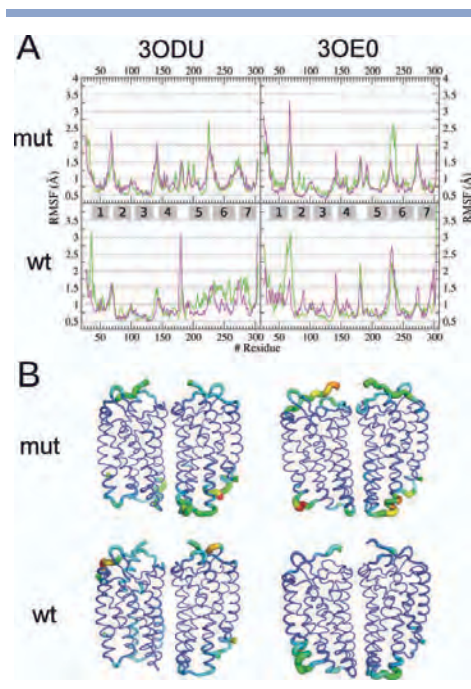


Figure 1

RMSF of the simulated (CXCR4)₂ systems. A: Green lines represent protomers A, and magenta lines account for protomers B. Location of TM helices are indicated with gray boxes. B: 3D representation of the RMSF, with residues' color increasing from blue to red as a function of the RMSF values, being the cartoon thickness also proportional to this property. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

With regards to the extracellular side, all the simulated systems maintained the initial hydrophobic contacts between the TM5 residues Leu194^{5,33}, Val197^{5,36}, Val198^{5,37}, Phe201^{5,40}, Met205^{5,44}, and Leu210^{5,49}, which validates their proposed role in the dimerization interface.⁸ Initial hydrogen bonds between protomers in this area involve symmetric residue pairs, that is, they occur between one residue of protomer A and another residue of protomer B, and vice versa (Table II and Fig. 4). Although originally proposed for the two crystal structures,⁸ we have identified that these hydrogen bonds are only present in the Xray (3ODU) structure, while in the corresponding assembly (3OE0) structure these residues are in proximity, but do not show an optimal hydrogen bond orientation. In any case, the simulations show that the hydrogen bond formed between residues Trp195^{5,34} of one protomer and Leu267^{6,63} of the opposite protomer is present, at least in one form, in all four of the systems. The hydrogen bond formed by Asn192^{5,31(IL2)} and

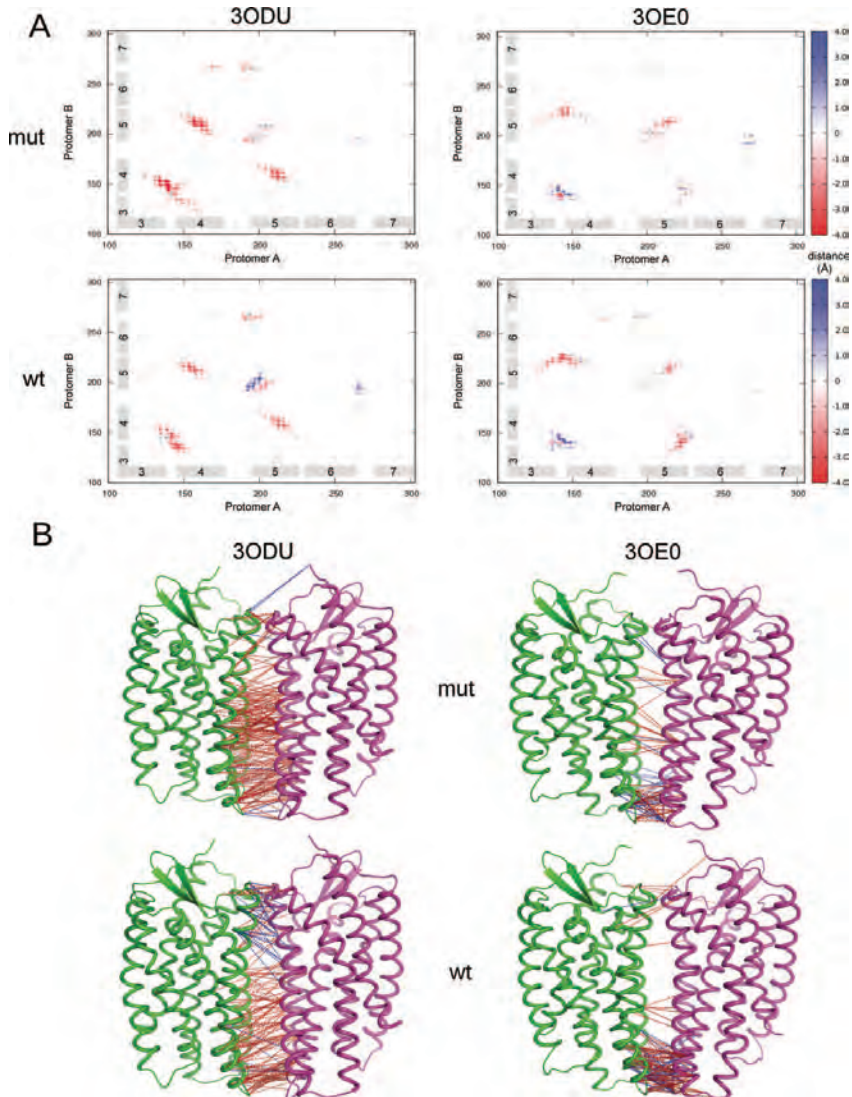


Figure 2

Contacts between protomers. Closing (red) and opening (blue) distances of the corresponding residue pairs are indicated as (A) Contact maps (location of TM helices indicated by gray boxes) and (B) in a 3D representation mapped on the corresponding starting structures of each (CXCR4)₂ model.

Glu268^{6.64} is also quite stable, and in (CXCR4)₂-Xray-wt it even occurs simultaneously in its two possible symmetrical forms. In addition, a new interaction involving

Asn192^{5.31(EL2)} and the main chain carbonyl of Leu267^{6.63} is observed in the (CXCR4)₂-assembly structures. Conversely, hydrogen bonds between residues Leu266^{6.62} and

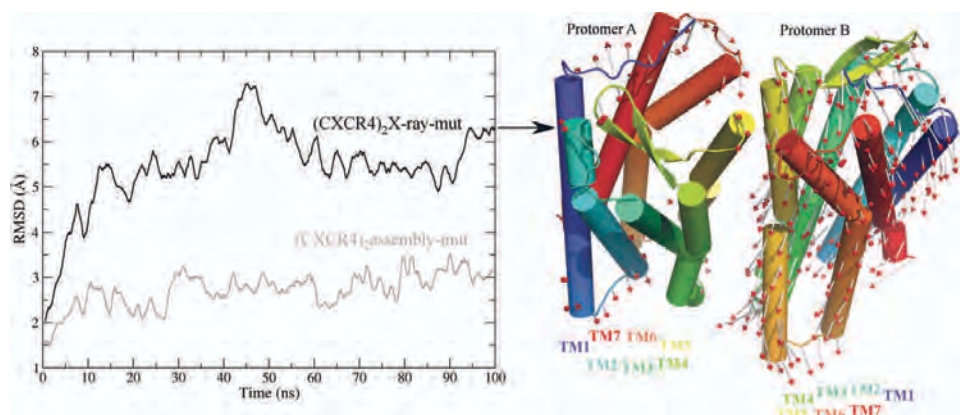


Figure 3

The backbone RMSD of protomer B, as calculated with all snapshots aligned to the backbone of protomer A (see also Table I). In the case of (CXCR4)₂-Xray-mut, the relative displacement of the protomers after the MD simulation is depicted with arrows on the initial 3D structure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Asn192^{5,31(EL2)} present insignificant occupancies except for (CXCR4)₂-Xray-wt, thus suggesting its irrelevance in the interaction between protomers despite being suggested by the crystal structure. In all cases, the relatively low mean number of simultaneous hydrogen bonds (between 2 and 4, see Table II), suggests that the dimerization is mainly stabilized by non-polar interactions, as it typically occurs for protein–protein interactions.³⁸

To evaluate the possible influence of the antagonist in the dimerization interface, an independent MD simulation was conducted on the (CXCR4-IT1t)₂ complex. Although the consideration of the two most probable protonation states of the ligand (i.e., with net charge +1 or +2, see Fig. S2), revealed no major effects in the MD simulations (data not shown), the later setup was selected for further analysis since it involves a salt bridge with the

Table II

Frequency (in Percentage of the MD Simulation Time) of Hydrogen Bonds Between Protomers

Residue pair		(CXCR4) ₂ -Xray			(CXCR4) ₂ -assembly	
Protomer A	Protomer B	mut	wt	Mut-(IT1t) ₂	mut	wt
Initial hydrogen bonds ^a						
Asn192 ^{5,31(EL2)}	Leu266 ^{6,62}	0.31%	9.80%	0.08%	0.00%	9.09%
Leu266 ^{6,62}	Asn192 ^{5,31(EL2)}	2.08%	61.61%	0.20%	0.01%	0.24%
Asn192 ^{5,31(EL2)}	Glu268 ^{6,64}	3.54%	35.23%	3.42%	0.01%	73.86%
Glu268 ^{6,64}	Asn192 ^{5,31(EL2)}	65.66%	53.28%	30.60%	1.52%	0.45%
Trp195 ^{5,34}	Leu267 ^{6,63}	0.66%	67.95%	0.34%	93.96%	91.80%
Leu267 ^{6,63}	Trp195 ^{5,34}	78.80%	2.01%	15.18%	1.81%	3.08%
New hydrogen bonds with >30% occupancy ^b						
Leu267 ^{6,63}	Asn192 ^{5,31(EL2)}	–	–	–	39.32%	48.07%
Tyr135 ^{3,51}	Tyr135 ^{3,51}	–	–	–	64.33%	10.93%
Ser225 ^{5,63}	Arg146 ^{4,35}	–	–	–	53.31%	35.59%
Ser228 ^{5,66}	Arg146 ^{4,35}	–	–	–	20.68%	35.47%
His140 ^{5,56(IL2)}	Tyr135 ^{3,51}	–	–	–	23.24%	59.47%
Average number of hydrogen bonds between subunits ^c		2.12 ± 0.83	3.76 ± 1.24	2.01 ± 1.10	3.93 ± 1.49	4.30 ± 1.24

^aAs observed in the initial structure of (CXCR4)₂-Xray dimer.

^bNew hydrogen bonds with >30% of occupancy are only found in 3OEO structures. For comparison between wild-type and mutant forms of (CXCR4)₂-assembly dimers, pairs with hydrogen bonds with occupancies below 30% are also indicated (in italics).

^cComputed for the second half of the production phase.

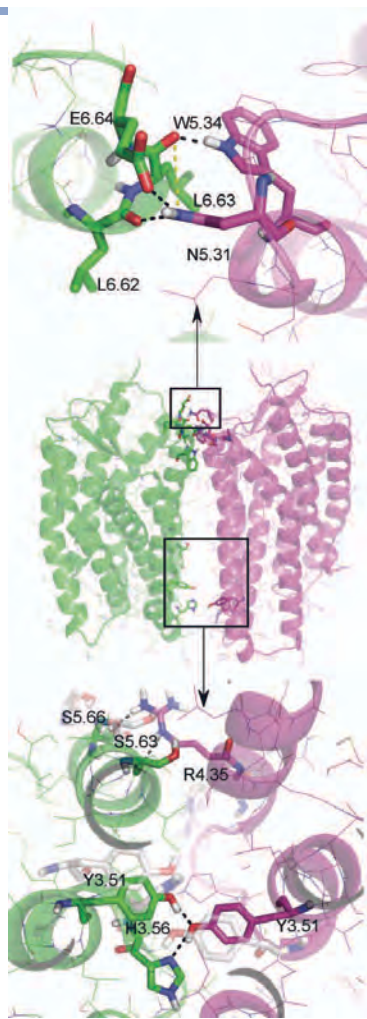


Figure 4

Hydrogen bonds between protomer A (green) and B (magenta), as collected in Table II. A detailed top view of hydrogen bonds at the extracellular side of the homodimer is on the top panel. The bottom panel depicts the novel hydrogen bond network observed in the intracellular side of the $(CXCR4)_2$ -assembly simulations, where the initial side chain conformations of the involved residues are represented in semitransparent white. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

negatively charged Glu288^{7,39}. In this case, a slight stabilizing effect of the ligand in the receptor dimerization interface (i.e., lower RMSD and smaller increase of BSA

between protomers) as compared to the corresponding apo simulation of $(CXCR4)_2$ -Xray-mut (see Table I) can be observed, while the average number of hydrogen bonds remains the same in the two cases, despite the lower frequency in certain hydrogen bond pairs (see Table II). Consequently, the ligand is not significantly altering the evolution of the dimerization interface but, as we will describe next, this system provides additional information about the network of interactions in the binding site.

Events involving key binding site residues

A region of the binding crevice that presents important conformational rearrangements in one of the simulated models is located around residue His113^{3,29}, with Asp171^{4,60}, and Arg188^{5,27(EL2)} in its vicinity (see Fig. 5). Comparing the two crystal structures of CXCR4 examined here (3ODU and 3OE0), the side chain of His113^{3,29} is found in a flipped state and different tautomeric form. This difference is likely due to the accommodation of the Arg2 side chain of the cyclic peptide CXV15 in the 3OE0 structure, which forms a salt bridge with Asp171^{4,60} and a hydrogen bond with His113^{3,29}, inducing a more open conformation of Arg188^{5,27(EL2)} as compared to the receptor in complex with IT1.⁸ Consequently, the corresponding tautomeric forms adopted by His113^{3,29}, supported by MolProbity predictions,²⁶ lead to a different set of polar interactions: in the 3ODU [$(CXCR4)_2$ -Xray] structure, the histidine coordinates the polar hydrogen in N δ 1 oriented to form a hydrogen bond with the main chain oxygen of Cys109^{3,25}. On the contrary, in 3OE0 [$(CXCR4)_2$ -assembly] structure, His113^{3,29} is preferably protonated in N ϵ 2 so an initial hydrogen bond with the main chain oxygen of Cys186^{5,25(EL2)} is achieved [see Fig. 5(A)]. This different initial configuration has an important effect in the salt bridge initially formed between Asp171^{4,60} and Arg188^{5,27(EL2)}. The interaction is maintained along the MD trajectories in the $(CXCR4)_2$ -Xray structures, with frequencies typically higher than 90% (see Fig. 5, panels B and D). However, in the simulations of the apo $(CXCR4)_2$ -assembly structures, the side chain of His113^{3,29} rotates to establish a stable hydrogen bond with Asp171^{4,60} (occupancy higher than 85%), significantly affecting the stability of the aforementioned salt bridge (reducing its frequency down to 25% and 63% for mutant and WT forms, respectively) (see Fig. 5, panels C and D). Mutational studies, together with the proximity of these residues in the starting structures, suggest an important role for this salt bridge in receptor function and/or ligand binding; mutants of both Asp171^{4,60} and Arg188^{5,27(EL2)} into residues without their respective original net charges have shown a marked loss of co-receptor function in HIV infection.^{39–41} In addition, mutation of Asp171^{4,60} into Asn produces an impaired binding of bi- and monocyclams, which are potent and selective CXCR4 antagonists with strong antiviral activity against HIV.^{41, 42} The simulations of the apo

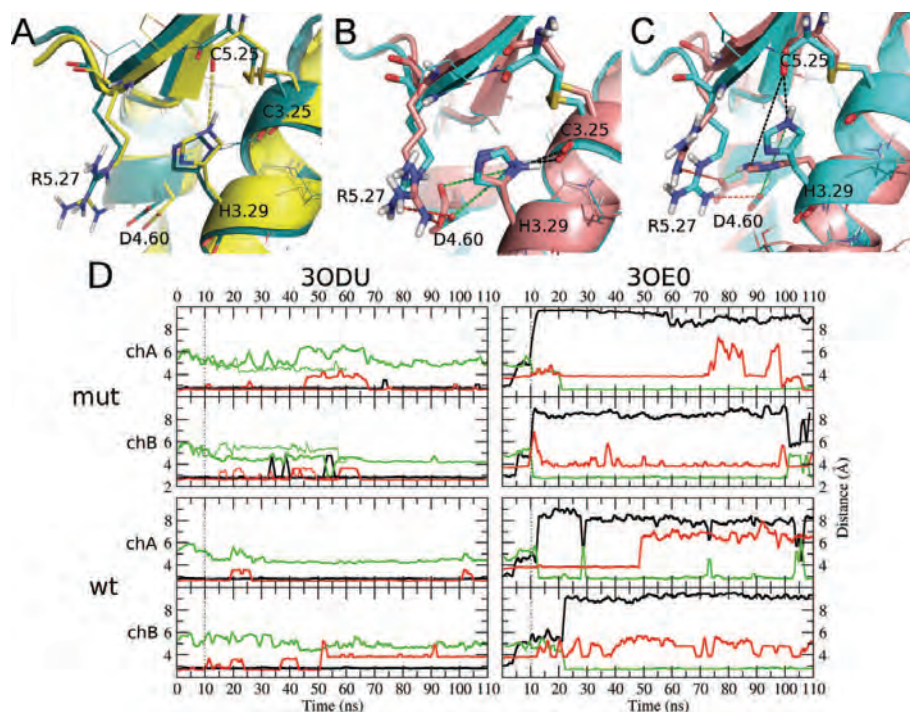


Figure 5

Configuration and dynamics of important binding site residues pointed out by mutagenesis data. **A:** Initial configurations of His113^{3,29} of 3ODU (green) and 3OE0 (yellow) structures. Initial (cyan) and final (brown) conformations of the cluster of residues for (CXCR4)₂-Xray (**B**) and (CXCR4)₂-assembly (**C**) simulations. **D:** Time evolution of the minimum distance between selected pairs of residues in the four systems; black line, His113^{3,29}/Cys109^{3,25} (3ODU, Xray), or His113^{3,29}/Cys186^{5,25(EL2)} (3OE0, assembly); green line, His113^{3,29}/Asp171^{4,60}; red line Asp171^{4,60}/Arg188^{5,27(EL2)}. The corresponding distances for the simulation of the complex (CXCR4-IT1t)₂-Xray-mut are shown in thin lines. Upper panels correspond to protomer A, and lower panels to protomer B. The vertical dotted line separates equilibration from production stages.

(CXCR4)₂-assembly system suggest that the rearrangement caused by the interaction of Asp171^{4,60} with Arg2 of the cyclic peptide in the parent crystal structure might be reverted when the ligand is not present, establishing residue interactions not previously observed in either of the two holo crystallographic structures of CXCR4. This observation is consistent with the idea that residues Asp171^{4,60} and Arg188^{5,27(EL2)} are individually important for binding of peptides and peptidomimetics, rather than the idea that a salt bridge between them would be key for ligand-independent function. On the contrary, it is important to note that in the simulation of the (CXCR4-IT1t)₂-Xray-mut system, the behavior of this region is analogous to the corresponding apo simulation [(CXCR4)₂-Xray-mut], as can be observed in Figure 5(D). This suggests that the organic molecule produces a negligible induced fit effect in these

residues upon binding, and that this particular receptor conformation is appropriate for evaluating the binding of small organic molecules exploring the same binding site region.

CONCLUSIONS

The use of all-atom MD simulations is becoming a routine technique to assess the increasing repertoire of GPCR crystal structures.⁴³ We report here all-atom MD simulations of different models of CXCR4 homodimers as revealed by X-ray crystallography. The results suggest that the dimerization model involving TM5 and TM6 is compatible with the explicit consideration of a cellular membrane, achieving increased surface contacts between protomers. Notably, no ligand binding is necessary to stabilize the observed dimeri-

zation interface, supporting experimental evidences of ligand-independent homodimerization of CXCR4 receptors.^{18,19} Indeed, the simulation (CXCR4-IT1t)₂-Xray-mut shows that the behavior of the dimerization interface is equivalent to that in the corresponding apo form. However, the orientation between protomers in the 3ODU structures is not at equilibrium, and an approach of the corresponding intracellular sides of the receptors accompanied by a relative rotation of one protomer results in closing distances between the corresponding TM4 regions of each protomer. Interestingly, the intracellular region between the two protomers was suggested to be filled by lipid molecules.⁸ However, our modeling suggests that no lipids should be present in this region, which is initially filled by a bubble of solvent molecules that subsequently dried along the MD simulations. Looking back to the parent crystal structures, one can appreciate that the dimerization interface of the 3ODU [(CXCR4)₂-Xray] presents a large contact surface between the two units of the T4L, which is fused on the intracellular side of each protomer. Consistent with this observation, the MD simulations presented here clearly indicate that the T4L, absent in our setup, biased the dimerization mode observed in this particular crystal structure. Indeed, the assembly model (3OE0), where the corresponding T4Ls do not interact with each other, is revealed as a more robust dimerization pose. The biological relevance of the dimerization interfaces achieved in the reported simulations, or the proposal of alternative models for CXCR4 homodimerization is an open question that deserves further studies, that is, performing BRET or FRET assays in conjunction with site-directed mutagenesis on selected residues to evaluate their influence in CXCR4 homodimerization. Importantly, the crystal structures of the μ - and κ -opioid receptors,^{44, 45} both released while this manuscript was under review, significantly contribute to our understanding of the GPCR oligomerization problem. In those structures, two dimerization interfaces are observed; one interface is located along the TM5-TM6 helices, quite similar to the CXCR4 dimerization model discussed here, and another one is between the TM1-TM2-TM8 helices. The two interfaces are highlighted as relevant for the formation of higher order oligomers.⁴⁴

Finally, our simulations suggest that the initial configuration of His113^{3,29} and Arg188^{5,27(EL2)} might be influenced by the corresponding ligand present in the crystal structures. Accordingly, in the structure crystallized with the small molecule (3ODU), the salt bridge between Asp171^{4,60} and Arg188^{5,27(EL2)} remains stable along the MD simulation time (including the control simulation in complex with the ligand), whereas in the (CXCR4)₂-assembly systems, this salt bridge is substituted by a strong hydrogen bond between His113^{3,29} and Asp171^{4,60}. This dynamic divergence exemplifies how important is the chemical nature of the co-crystallized ligands for the selection of a representative receptor conformation, which would be further used as a template structure for homology modeling of related receptors, or as a crucial starting point for SBDD efforts.

ACKNOWLEDGMENTS

Computations were performed at CESGA, initially assisted by Leonardo Mokarzel-Falcón. D.R. is recipient of a predoctoral grant from the Fondo de Investigación en Salud (ISCIII). H.G.T. is a researcher of the Isidro Parga Pondal program (Xunta de Galicia, Spain). The authors are thankful to Dr. Vsevolod Katritch for his critical reading of the manuscript and to Angela Walker for proofreading the final text.

REFERENCES

1. Kenakin T. Principles: receptor theory in pharmacology. *Trends Pharmacol Sci* 2004;25:186–192.
2. Viola A, Luster AD. Chemokines and their receptors: drug targets in immunity and inflammation. *Annu Rev Pharmacol Toxicol* 2008;48:171–197.
3. Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* 1996;272:872–877.
4. Brelot A, Heveker N, Montes M, Alizon M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* 2000;275:23736–23744.
5. Berger EA, Murphy PM, Farber JM. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* 1999;17:657–700.
6. Balkwill F. The significance of cancer cell expression of the chemokine receptor CXCR4. *Semin Cancer Biol* 2004;14:171–179.
7. Hanson MA, Stevens RC. Discovery of new GPCR biology: one receptor structure at a time. *Structure* 2009;17:8–14.
8. Wu B, Chien EY, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, Hamel DJ, Kuhn P, Handel TM, Cherezov V, Stevens RC. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 2010;330:1066–1071.
9. Selent J, Kaczor AA. Oligomerization of G protein-coupled receptors: computational methods. *Curr Med Chem* 2011;18:4588–4605.
10. Fotiadis D, Liang Y, Filipek S, Saperstein DA, Engel A, Palczewski K. Atomic-force microscopy: rhodopsin dimers in native disc membranes. *Nature* 2003;421:127–128.
11. Liang Y, Fotiadis D, Filipek S, Saperstein DA, Palczewski K, Engel A. Organization of the G protein-coupled receptors rhodopsin and opsin in native membranes. *J Biol Chem* 2003;278:21655–21662.
12. Fanelli F, De Benedetti PG. Inactive and active states and supramolecular organization of GPCRs: insights from computational modeling. *J Comput Aided Mol Des* 2006;20:449–461.
13. Filizola M, Weinstein H. The study of G-protein coupled receptor oligomerization with computational modeling and bioinformatics. *FEBS J* 2005;272:2926–2938.
14. Percherancier Y, Berchiche YA, Slight I, Volkmer-Engert R, Tamamura H, Fujii N, Bouvier M, Heveker N. Bioluminescence resonance energy transfer reveals ligand-induced conformational changes in CXCR4 homo- and heterodimers. *J Biol Chem* 2005;280:9895–9903.
15. Dean MK, Higgs C, Smith RE, Bywater RP, Snell CR, Scott PD, Upton GJ, Howe TJ, Reynolds CA. Dimerization of G-protein-coupled receptors. *J Med Chem* 2001;44:4595–4614.
16. George SR, O'Dowd BF, Lee SR. G-protein-coupled receptor oligomerization and its potential for drug discovery. *Nat Rev Drug Discovery* 2002;1:808–820.
17. Wang J, Norcross M. Dimerization of chemokine receptors in living cells: key to receptor function and novel targets for therapy. *Drug Discov Today* 2008;13:625–632.

18. Wang J, He L, Combs CA, Roderiquez G, Norcross MA. Dimerization of CXCR4 in living malignant cells: control of cell migration by a synthetic peptide that reduces homologous CXCR4 interactions. *Mol Cancer Ther* 2006;5:2474–2483.
19. Babcock GJ, Farzan M, Sodroski J. Ligand-independent dimerization of CXCR4, a principal HIV-1 coreceptor. *J Biol Chem* 2003; 278:3378–3385.
20. Choi WT, Duggineni S, Xu Y, Huang Z, An J. Drug Discovery Research Targeting the CXC Chemokine Receptor 4 (CXCR4). *J Med Chem* 2011.
21. Rodríguez D, Piñero A, Gutiérrez-de-Terán H. Molecular dynamics simulations reveal insights into key structural elements of adenosine receptors. *Biochemistry* 2011;50:4194–4208.
22. Thoma G, Streiff MB, Kovarik J, Glickman F, Wagner T, Beerli C, Zerwes HG. Orally bioavailable isothioureas block function of the chemokine receptor CXCR4 in vitro and in vivo. *J Med Chem* 2008;51:7915–7920.
23. Schrödinger LLC. Macromodel, version9.7. New York, NY; 2009.
24. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
25. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci* 2006;15:2507–2524.
26. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* 2007;35:W375–W383.
27. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res* 2004;32:W665–W667.
28. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 2008;4:435–447.
29. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 2001;105:6474–6487.
30. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B, editor. *Intermolecular forces*. Dordrecht: D. Reidel Publishing Company; 1981. pp 331–342.
31. Chakrabarti N, Neale C, Payandeh J, Pai EF, Pomès R. An iris-like mechanism of pore dilation in the CorA magnesium transport system. *Biophys J* 2010;98:784–792.
32. Berger O, Edholm O, Jähnig F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys J* 1997;72: 2002–2013.
33. Nose S, Klein ML. Constant pressure molecular-dynamics for molecular-systems. *Mol Phys* 1983;50:1055–1076.
34. Darden T, York D, Pedersen L. Particle mesh Ewald: an $N \log(N)$ method for Ewald sums in large systems. *J Chem Phys* 1993;98: 10089–10089.
35. Marvin 5.7.0; 2011. ChemAxon (<http://www.chemaxon.org>).
36. Díaz L, Bujons J, Delgado A, Gutiérrez-de-Terán H, Aqvist J. Computational prediction of structure-activity relationships for the binding of aminocyclitols to beta-glucocerebrosidase. *J Chem Inf Model* 2011;51:601–611.
37. Ballesteros JA, Weinstein H. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods neuroscience*. San Diego: Academic Press; 1995. pp 366–428.
38. Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. *Q Rev Biophys* 2008;41:133–180.
39. Wang ZX, Berson JF, Zhang TY, Cen YH, Sun Y, Sharron M, Lu ZH, Peiper SC. CXCR4 sequences involved in coreceptor determination of human immunodeficiency virus type-1 tropism. Unmasking of activity with M-tropic Env glycoproteins. *J Biol Chem* 1998;273: 15007–15015.
40. Chabot DJ, Zhang PF, Quinnan GV, Broder CC. Mutagenesis of CXCR4 identifies important domains for human immunodeficiency virus type 1 X4 isolate envelope-mediated membrane fusion and virus entry and reveals cryptic coreceptor activity for R5 isolates. *J Virol* 1999;73:6598–6609.
41. Hatse S, Princen K, Gerlach LO, Bridger G, Henson G, De Clercq E, Schwartz TW, Schols D. Mutation of Asp(171) and Asp(262) of the chemokine receptor CXCR4 impairs its coreceptor function for human immunodeficiency virus-1 entry and abrogates the antagonistic activity of AMD3100. *Mol Pharmacol* 2001;60:164–173.
42. Gerlach LO, Skerlj RT, Bridger GJ, Schwartz TW. Molecular interactions of cyclam and bicyclam non-peptide antagonists with the CXCR4 chemokine receptor. *J Biol Chem* 2001;276:14153–14160.
43. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, Arlow DH, Rasmussen SG, Choi HJ, Devree BT, Sunahara RK, Chae PS, Gellman SH, Dror RO, Shaw DE, Weis WI, Caffrey M, Gmeiner P, Kobilka BK. Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 2011;469:236–240.
44. Manglik A, Kruse AC, Kobilka TS, Thian FS, Mathiesen JM, Sunahara RK, Pardo L, Weis WI, Kobilka BK, Granier S. Crystal structure of the mu-opioid receptor bound to a morphinan antagonist. *Nature* 2012.
45. Wu H, Wacker D, Mileni M, Katritch V, Han GW, Vardy E, Liu W, Thompson AA, Huang XP, Carroll FI, Mascarella SW, Westkaemper RB, Mosier PD, Roth BL, Cherezov V, Stevens RC. Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 2012.

Annex I

Rodríguez D. and Gutiérrez-de-Terán H.

Adenosine Receptors: a systematic study of ligand binding based on the crystal structure of hA2A receptor.

Poster presentation at the *International Workshop In Memoriam of Ángel Ortiz*. Madrid, January 26-28 (**2009**).

ADENOSINE RECEPTORS: A SYSTEMATIC STUDY OF LIGAND BINDING BASED ON THE CRYSTAL STRUCTURE OF hA_{2A} RECEPTOR



David Rodríguez, Hugo Gutiérrez-de-Terán *
 Fundación Pública Galega de Medicina Xenómica – SERGAS
 Complejo Hospitalario Universitario de Santiago, E-15706 Santiago de Compostela (Spain)



* hugo.teran@usc.es

INTRODUCTION

Adenosine receptors (ARs) are a family of membrane proteins belonging to the superfamily of 7TM G-protein coupled receptors (GPCRs). Four AR subtypes (A₁, A_{2A}, A_{2B} and A₃) have been cloned for different species, and they receive much attention as potential drug targets in disorders such as schizophrenia, Parkinson, cardiovascular diseases, inflammatory processes or asthma. Even though high sequence identity (82-93%) exists between all subtypes, there are remarkable differences in ligand binding affinity within the family [Fig. 1].

Until very recently, the different structural studies about this family of receptors were based on homology models, mainly using the bovine rhodopsin as template, or either ligand based techniques [1]. The lately published structure of the hA_{2A} receptor in complex with the potent antagonist ZM241385 (PDB code 3EML) [2] opened the door to a more confident structural analysis of important issues such as affinity and selectivity among ARs.

METHODS

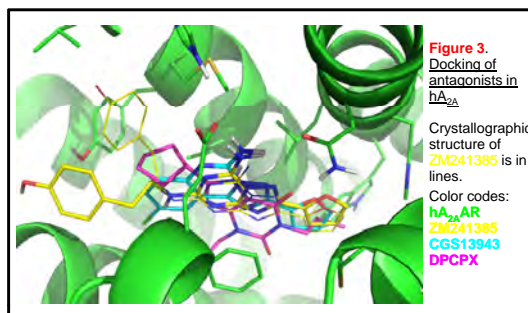
- 1) Sequence and pseudo-sequence alignment of ARs.** Full alignment with ClustalX. Pseudo-sequences of binding sites were determined by selecting residues within 5Å of ZM241385 [Fig. 2].
- 2) Homology modeling of human A₁, A_{2B} and A₃ ARs using hA_{2A} as template (3EML), performed with MODELLER 9v4 [3], followed by evaluation and refinement with Molprobit [4].**
- 3) Automated docking of every antagonist / AR combination with GOLD [5].**

RESULTS

→ **Docking on A_{2A}:** solutions overlapped with crystallographic antagonist binding mode. **Biological superimposition of the three ligands is in agreement with previous QSAR studies [6] [Fig. 3].**

→ **Comparison of A_{2B} / A_{2B} docking: analogous superimposition was found for both receptors.** Slight differences in affinity could be attributed to **residue 6.51 [Fig. 2].**

→ **Docking of ZM241385 against A₁ and A₃.** Inverted pose was found for A₁ A₃ binding site is narrower and deeper compared to A_{2A}. **These evidences could justify their low affinity [Fig. 4].**



FUTURE WORK

- 1) Molecular Dynamics Simulations,** including binding free energy prediction, will be performed to optimize ligand-receptor interactions and provide a rationale of the binding affinity [7].
- 2) Proposal of site-directed mutagenesis of the residues determined to be crucial by the previous studies.**

AIM OF THE STUDY

- 1) To perform a compared analysis of the binding site of the different ARs on the basis of the new hA_{2A} structure.**
- 2) To elucidate the binding mode of the different inhibitor scaffolds for ARs, and to determine their biological superimposition.**

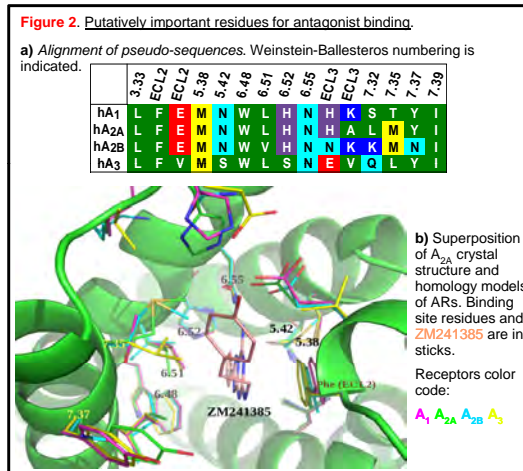
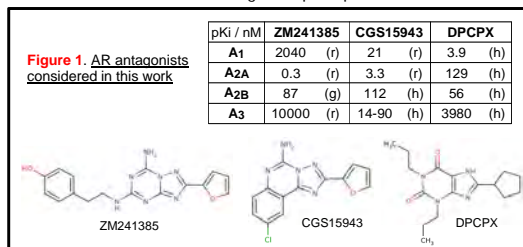
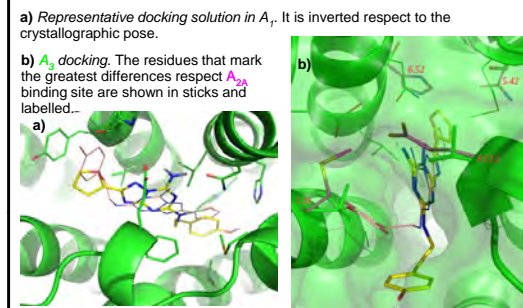


Figure 4. ZM241385 binding mode in human A₁ and A₃. Side chains of pseudo-sequence amino acids are represented. Color codes for ZM241385: docking solutions (green), crystallographic structure (yellow).



[1] Moro S et al. (2005) Trends in Pharmacol Sci 26:44-51

[2] Jaankola VP et al. (2008) Science 322(5905):1211-7

[3] Sali A et al. (2003) J. Mol. Biol. 234, 779-815

[4] Davis IW et al. (2007) Nucleic Acids Res. 35:W375-W383

[5] Jones G et al. (1997) J. Mol. Biol. 272:7-48

[6] Gutiérrez-de-Terán H et al. (2003) Presentation at SEGT meeting, Sevilla.

[7] Gutiérrez-de-Terán H et al. (2004) ChemBioChem 5:841-849a

Annex II

Rodríguez, D.; Sotelo, E.; Bender, A.; and Gutiérrez-de-Terán H.

Discovery of potent and selective adenosine receptor ligands via multi-objective design

Abstract of the oral presentation given at the *243rd ACS National Meeting & Exposition*. San Diego (California), March 25-29 (**2012**)

The four Adenosine Receptors (ARs), namely A₁, A_{2A}, A_{2B} and A₃, are a family of GPCRs with a well recognized therapeutic potential in disorders such as inflammatory processes, asthma or Parkinson disease.

We present results of a ligand-based virtual screening (LBVS) aimed to obtain novel chemotypes with enough potency and selectivity profiles to be considered lead compounds on each of the four human ARs. Multi-objective selection is performed with the molecular similarity searching tool MOLPRINT 2D, employing curated datasets from the ChEMBL database in order to develop bioactivity models for each AR. These models are then used to screen both public and proprietary databases looking for selective novel compounds towards these receptors.

The virtual hits are being currently pharmacologically evaluated for their validation in the four ARs. A comparison of the outcome of this study with recent structure-based VS on ARs is presented.

Annex III

SUMMARY IN SPANISH RESUMEN EN CASTELLANO*

Tesis doctoral:

“Métodos Computacionales para la Caracterización Estructural y Dinámica de Receptores Acoplados a Proteína-G: Aplicaciones en el Diseño de Fármacos”

* Las siglas indicadas en este resumen se derivan de los términos correspondientes en lengua inglesa, en concordancia con el texto principal.

En la actualidad, la industria farmacéutica se enfrenta a un escenario complejo en la investigación y desarrollo de nuevos fármacos. Los crecientes costes y las dificultades en la obtención de nuevas entidades químicas exigen un rendimiento óptimo en las diferentes etapas necesarias para la aprobación de un medicamento. Las herramientas computacionales están ampliamente integradas en el proceso de desarrollo de nuevos fármacos, tanto en el ámbito académico como en la industria, y su aplicación abarca varias con aplicaciones como *i*) la realización de cribados virtuales (VS) de grandes quimiotecas, enriqueciendo posteriores ensayos farmacológicos de alto rendimiento sobre dianas de interés, o *ii*) el diseño racional en la optimización de compuestos cabezas de serie (*leads*) usando métodos basados en la estructura (SB) del receptor tales como predicciones acoplamiento molecular (*docking*) o simulaciones de dinámica molecular (MD).

En la presente tesis, se ha empleado una amplia gama de técnicas computacionales para el estudio de varios miembros de la superfamilia de receptores acoplados a proteína-G (GPCRs), un grupo de proteínas de membrana de gran interés farmacológico. El principal afán ha sido la adecuada integración de métodos computacionales de diseño de fármacos basado en receptor (SB) y el ligando (LB), junto con el empleo de técnicas de modelado por homología para la caracterización estructural, y de simulación de MD para el estudio conformacional de estos receptores, un aspecto clave en su función.

Los GPCRs constituyen la principal superfamilia de proteínas integrales de membrana, cuya función es la transducción de señales desde el medio extracelular hacia el citoplasma. El mecanismo canónico de señalización de los GPCRs implica la activación de la proteína-G heterotrímica intracelular, mediante la disociación de su subunidad α . Existen diferentes tipos de proteínas-G, cada una estimulando específicas rutas de señalización bioquímicas. Gracias a esta diversidad de señalización, los GPCRs median una gran variedad de funciones fisiológicas, en parte explicando su condición de principal diana farmacológica, ya que al menos el 30% de los fármacos comercializados a día de hoy se dirigen a un GPCR. Los GPCRs son receptores muy flexibles, explorando diferentes conformaciones cada una con su correspondiente implicación funcional, abarcando desde el estado inactivo al activo. Los ligandos extracelulares pueden influir en este equilibrio a fin de inducir diferentes respuestas biológicas, produciendo un aumento (agonistas completos y parciales), mantenimiento (antagonistas) o reducción (agonista inverso) de la respuesta biológica basal (sin ligando) del receptor.

La mayoría de las familias de GPCRs se caracterizan por unirse a un ligando específico, el agonista natural de un receptor dado. Típicamente, los ligandos de GPCRs se unen al sitio ortostérico, el mismo del agonista natural, situado en un bolsillo profundo e hidrofóbico en la región extracelular del receptor. La gran diversidad química de los ligandos naturales de GPCRs está en línea con la importante variedad de funciones fisiológicas mediadas por estos receptores. Atendiendo al tipo de agonistas naturales y al árbol filogenético basado en su secuencia, los 900 GPCRs humanos se dividen en 5 grupos, del cual el clan A (o rama de receptores similares a rodopsina) es el más poblado y el que cuenta con un mayor interés farmacológico. Los receptores estudiados en la siguiente tesis pertenecen a dicho clan, y se describen en más detalle en los siguientes párrafos:

RESUMEN

i) Receptores involucrados en esquizofrenia. La esquizofrenia es una enfermedad psiquiátrica que afecta significativamente a la calidad de vida de aproximadamente un 1% de la población mundial. A pesar de los constantes avances en la caracterización bioquímica y farmacológica de la misma, los fármacos que se emplean para su tratamiento cuentan con numerosas limitaciones terapéuticas y contraindicaciones. De hecho, casi 50 años después de su descubrimiento, la clozapina se considera el antipsicótico de referencia, a pesar de contar con serios efectos secundarios. La dificultad en el tratamiento de la esquizofrenia deriva de la necesidad de alcanzar la apropiada señalización dopaminérgica en regiones específicas del cerebro, mediante un complejo equilibrio de actividades farmacológicas en un amplio rango de receptores. La mayoría de estas dianas son GPCRs, siendo clave el de serotonina 5-HT_{2A}, y el de dopamina D₂, además de familias de los receptores aminérgicos como los α y β adrenérgicos (ADR), muscarínicos o de histamina. Los antipsicóticos desarrollados recientemente presentan esqueletos químicos similares a la clozapina, por lo que el descubrimiento de moléculas novedosas con capacidad de obtener el perfil farmacológico deseado proporcionaría valiosas herramientas farmacológicas para tratar esta esquiiva enfermedad.

ii) Receptores de adenosina. El nucleósido adenosina regula un amplio rango de funciones biológicas a través de la activación de los receptores de adenosina (ARs). Éstos se subdividen en cuatro subtipos: A₁, A_{2A}, A_{2B} y A₃. Los ARs son dianas farmacológicas altamente relevantes debido a su extenso potencial en una amplia variedad de patologías. Su modulación presenta interesantes indicaciones, y los esfuerzos de diseño de fármacos se han centrado tradicionalmente en el descubrimiento de ligandos ortostérico químicamente similares respecto a productos naturales que se unen a ARs, tanto agonistas (adenosina) como antagonistas (derivados xantínicos tales como la cafeína). Sin embargo, el desarrollo de fármacos novedosos potentes y selectivos de esta familia es todo un reto debido a la similitud estructural de los ARs humanos.

iii) Receptores de quimioquina. Las quimioquinas son proteínas de pequeño tamaño que inducen la quimiotaxis de células somáticas mediante la activación de sus receptores naturales. Hasta 20 miembros de los receptores de quimioquinas se han descrito en humanos. Se asocian con procesos patologías tales como la inflamación y las enfermedades autoinmunes. Uno de los miembros más estudiado de esta familia es el receptor CXC 4 (CXCR4), que participa en el proceso de entrada en la célula huésped por parte del virus VIH-1. Además, CXCR4 se ha implicado en el desarrollo de más de 20 tipos de cáncer, ofreciendo así una amplia gama de posibles aplicaciones terapéuticas.

La estructura tridimensional de alta resolución (típicamente por debajo de 3 Å) de una proteína resuelta mediante técnicas experimentales —como la cristalografía de Rayos-X— supone una valiosísima fuente de información para el diseño de fármacos. Sobre estas estructuras se pueden emplear de técnicas computacionales basadas en estructura del receptor (SB) para predecir la forma de unión de ligandos, y posibles optimizaciones de los mismos. Cuando no existe dicha estructura, se pueden utilizar alternativamente métodos basados en ligando (LB), que relacionan las variaciones de estructura y actividad de grupos de compuestos químicos. Finalmente, también existen métodos de predicción de la estructura del receptor como el modelado por homología. En lo relativo a los GPCRs, la obtención de estructuras de Rayos-X es una tarea extremadamente compleja, ya que son receptores integrales de membrana con flexibilidad inherente. Sin embargo,

tradicionalmente se ha reconocido su topología global, formada por 7 hélices transmembrana (TM) conectadas por 3 lazos extracelulares y 3 intracelulares mediante distintos métodos biofísicos. Hasta el año 2007 no se desarrollaron revolucionarias metodologías en el área de la cristalografía de GPCRs, permitiendo la resolución de hasta 14 miembros de distintas familias de estos receptores en conformaciones de diferente significación funcional (unidas a agonistas, antagonistas o agonistas inversos). A pesar de estos avances, con un gran impacto en la biología estructural y en el diseño de fármacos mediante técnicas computacionales, la cobertura estructural de toda la superfamilia es aún limitada. De hecho, numerosos receptores de gran interés farmacológico, como el de de oxitocina o los purinérgicos P2Y siguen sin estar caracterizados experimentalmente. De ahí se desprende la gran demanda existente para obtener modelos computacionales de calidad en una gran cantidad de GPCRs.

La resolución de las estructuras cristalográficas de GPCRs ha revelado la importancia de pequeños motivos estructurales en la función de estos receptores, junto con distintas exploraciones computacionales de los mismos. Cabe destacar *i)* el *ionic lock* (“cerradura iónica”) formado por un puente salino entre el residuo conservado Arg^{3.50*} y con un Asp/Glu^{6.30}, localizados en la zona intracelular de TM3 y TM6. Esta interacción se ha sido tradicionalmente relacionada con un estado inactivo del receptor. Por su parte, *ii)* el cambio del rotámero del residuo Trp^{6.48}, o *toggle switch* (“interruptor de palanca”) —localizado en el extremo inferior del sitio de unión de ligandos de GPCRs— se asocia con su activación. Asimismo, otros motivos estructurales también son descritos y analizados en el contexto de esta tesis.

En cuanto a la topología extracelular de los GPCRs, las estructuras cristalográficas resueltas muestran una gran variabilidad según la familia del receptor. Ésta puede considerarse la puerta de entrada del ligando al sitio de unión. Los tres distintos tipos de receptores estudiados en esta tesis (serotonina, ARs y receptores de quimioquinas) se benefician de la resolución de estructuras cristalográficas de β 2ADR, A_{2A}AR y CXCR4 unidos a antagonistas, que muestran los residuos clave para la unión de ligandos. Asimismo, nuevas evidencias experimentales de la dimerización de GPCRs, un aspecto de gran importancia en la función de estos receptores, han sido estudiadas mediante métodos computacionales en esta tesis.

En cuanto a las técnicas empleadas en los trabajos aquí presentados, se pueden dividir en tres grandes bloques: modelado por homología (para la predicción computacional de la estructuras de un receptor), métodos de diseño de fármacos (basados en estructura del receptor o en ligando) y simulaciones de dinámica molecular:

i) Modelado por homología. A pesar de la elevada diversidad de secuencia entre los GPCRs, se pueden obtener modelos computacionales de gran fiabilidad de estos receptores, gracias a la topología de 7TM conservada en la superfamilia. Las nuevas estructuras experimentales proveen moldes de creciente fiabilidad para el modelado de un

* Los residuos son numerados de acuerdo con el patrón X.YY, donde la X denota la hélice transmembrana, e YY es un número correlativo siguiendo la secuencia del receptor, asignando el 50 al residuo más conservado en GPCRs en la hélice correspondiente.

RESUMEN

mayor número de receptores de esta superfamilia. Para la realización de modelos por homología, se empleó el programa MODELLER, además de otras herramientas para evaluar y refinar dichos modelos estructurales.

ii) Métodos de diseño de fármacos asistido por ordenador. Como ya se introdujo anteriormente, pueden ser SB o LB. Respecto a los SB, estudios de *docking* fueron realizados con el programa GOLD, el cual predice posibles formas de unión de un ligando mediante una búsqueda conformacional con un algoritmo genético, y una función de puntuación empírica (relacionada con la energía libre de unión ligando-receptor). En cuanto a los métodos LB, estudian grupos de ligandos y sus respectivas actividades biológicas en conjunto, para explicar relaciones de estructura-actividad (SAR) o para obtener nuevas entidades químicas con características deseadas. Siguen el principio de similitud molecular, por el cual compuestos similares deben presentar características parecidas. Los diferentes acercamientos LB emplean distintas descripciones matemáticas de las moléculas (2D ó 3D) y posteriores medidas métricas de su similitud química. En la presente tesis se han utilizado métodos de VS basado en forma de ligando (con la suite de programas OpenEye), relaciones de estructura-actividad cuantitativas (3D-QSAR, con el programa Pentacle), y búsqueda de similitud basada en ambientes atómicos (con la herramienta MOLPRINT2D). Asimismo, combinaciones complementarias de métodos LB y SB se han explorado en esta tesis.

iii) Simulaciones de MD. El objetivo principal de esta técnica es reproducir la evolución dinámica de la estructura de biomoléculas para extraer conclusiones funcionales. Dichas simulaciones se basan en una descripción de mecánica molecular del sistema, mediante un campo de fuerza o *force field*, que evalúa las distintas interacciones entre átomos para calcular la energía potencial del sistema. Posteriormente, el cálculo de las sucesivas coordenadas de los átomos a lo largo del tiempo sigue las leyes de Newton. Gracias a los avances en programación y equipos de supercomputación en esta área, la precisión y las escalas de tiempo accesibles son cada vez mayores.

La principal motivación de esta tesis es proporcionar una metodología computacional para la caracterización de la estructura y dinámica de los miembros seleccionados de GPCRs, dirigidas al esclarecimiento de los mecanismos de activación de miembros de estos receptores, así como la realización de proyectos de descubrimiento y diseño de fármacos. Los resultados obtenidos pueden ser englobados dentro de los siguientes objetivos:

1) Desarrollo y aplicación de una metodología robusta para el modelado por homología de GPCRs, con aplicación en el contexto del diseño de fármacos asistido por ordenador.

En esta tesis se ha realizado modelado por homología de GPCRs con el programa MODELLER, por el cual se emplea una estructura (experimental) como molde, y se deriva la estructura deseada a partir de un alineamiento de secuencia previamente realizado, en este caso con ClustalX2. Además, se desarrolló un protocolo integral incluyendo etapas de refinado y validación de los modelos estructurales.

El anteriormente citado protocolo de modelado computacional se empleó en un primer estudio donde se descubrieron nuevos esqueletos químicos como potenciales antipsicóticos de una quimioteca original de 1.622 compuestos, mediante un VS basado en

forma de ligando. En este caso se empleó la clozapina como compuesto de referencia para el VS, descubriéndose *hits* farmacológicos con esqueletos purínicos en el rango μM sobre el receptor 5-HT_{2A}. Posteriormente se generó un modelo computacional de dicho receptor con la estructura de β2ADR como molde, donde las formas de unión predichas de clozapina y el hit más activo se superponían de forma análoga respecto al método de VS. El apartado SB proporcionó razones estructurales de la relativamente reducida afinidad de los *hits*.

Una posterior evaluación de la metodología desarrollada en esta tesis tuvo lugar con la participación en el concurso GPCR Dock 2008, donde se sugería a la comunidad científica a predecir computacionalmente la estructura del receptor A_{2A}AR unida al potente antagonista ZM241385 justo antes de hacerse pública su estructura cristalográfica. Empleando de nuevo β2ADR como molde, se generaron 5 modelos por homología del receptor, con las correspondientes formas de unión del compuesto problema. Nuestras predicciones se encuentran por encima de la media, destacando la precisión de los modelos del receptor. Sin embargo, la forma de unión de ZM241385 predicha por nuestro grupo de investigación se encontraba invertida respecto a la cristalográfica, ejemplificando la importancia de incorporar adecuadamente la máxima información experimental posible.

Tras la publicación de la estructura del receptor A_{2A}AR, se realizaron modelos por homología de los otros tres subtipos de adenosina humanos (A₁, A_{2B} y A₃). Se caracterizaron las principales diferencias estructuralmente en el sitio de unión, además de realizar un estudio sistemático de *docking* de antagonistas de distintas familias, con distintos perfiles de selectividad. Las diferencias en las formas de unión de dichos ligandos están en línea con los datos experimentales de afinidad, apoyando la metodología empleada.

Sobre la base de este estudio de modelización de los miembros de la familia de ARs, nuestro laboratorio contribuyó diseño computacional y a la racionalización del perfil farmacológico de una novedosa quimiteca de diaril-pirimidinas, potentes y selectivas sobre el receptor A₃AR. Se realizó el *docking* sistemático de los compuestos sobre el anterior modelo de A₃AR, obteniendo un modo de unión consenso, posteriormente empleado para generar un modelo 3D-QSAR con una calidad estadística satisfactoria. De este modo se extrajeron las principales características farmacofóricas de la serie, sugiriendo una relación recíproca entre el tamaño de los sustituyentes de los ligandos explorando distintos subsitios lipofílicos del receptor. Una ventaja de esta combinación de metodologías SB y LB es que las descripciones 3D-QSAR se pueden proyectar hacia el sitio de unión, explicando la reducida afinidad de una de las sub-series de compuestos. Además, claves sobre la selectividad de los compuestos pueden detectarse a través del análisis de secuencias de sitios de unión de ARs descritos anteriormente, resaltando sustituciones en las posiciones 5.30, 5.42, 6.52 y 7.35 en este receptor respecto a los otros subtipos de ARs. El protocolo computacional se empleó en el diseño de un conjunto de 6 nuevos ligandos, con excelentes resultados.

La última combinación de métodos SB y LB ha sido realizada en el contexto de un proyecto en progreso, para el descubrimiento de nuevos compuestos novedosos, potentes y selectivos sobre los subtipos de ARs. La quimiteca de la Red Gallega de Descubrimiento de Fármacos fue inicialmente cribada con modelos de bioactividad generados con MOLPRINT2D, los cuales fueron entrenados con datos de actividad de compuestos anotados en la base de datos ChEMBL. Un posterior filtrado por criterios fisicoquímicos y

RESUMEN

de similitud fue realizado antes de la última etapa en la que se evaluaron las formas de unión de los compuestos en modelos por homología de los 4 ARs. De los 40 compuestos ensayados farmacológicamente, se obtuvieron 11 *hits*, dos de ellos en el rango sub- μ M en los subtipos A₁AR y A_{2B}AR.

2) Estudio de la dinámica y el equilibrio conformacional de GPCRs.

Para alcanzar este objetivo, en el contexto de esta tesis se elaboró un protocolo para configurar simulaciones de dinámica molecular de GPCRs, con un modelo atómico de la membrana y solvente explícito, empleando herramientas del programa GROMACS. Mediante esta metodología se publicaron dos estudios de dinámica molecular:

- Caracterización del equilibrio conformacional de receptores de adenosina. Partiendo de la estructura cristalográfica del receptor A_{2A}AR, y con el modelo por homología del receptor A_{2B}AR generado previamente, se realizaron simulaciones de MD en diversas condiciones, con extensiones en el rango de los 100 nanosegundos. Los principales eventos estudiados se relacionan con los principales motivos estructurales de GPCRs introducidos anteriormente: se observó una reformación del *ionic lock* (se hipotetiza que su estado inicial está influenciado por el método de cristalización empleado), y el cambio conformacional del *toggle switch* hacia una conformación putativamente más activa (soportado por cambios conformacionales concertados de residuos contiguos). Además, se estudió la dinámica de interacciones entre residuos conservados en ARs, ofreciendo explicaciones sobre el equilibrio conformacional de estas regiones en el contexto de la unión de ligandos.

- El protocolo computacional presentado anteriormente fue adaptado para el caso de dímeros de GPCRs, estudiándose la estabilidad de dos de las recientes estructuras cristalográficas de CXCR4 mediante simulaciones de MD, en el mismo rango de tiempo que las anteriores. El modo de dimerización observado permaneció estable, sobre todo para el caso de la estructura unida al derivado peptídico CVX15, por lo que este estudio computacional soporta la relevancia biológica de las interacciones observadas, en concordancia con más recientes estructuras cristalográficas de receptores opioides. Asimismo, también se evaluó la dinámica de una región del sitio de unión de CXCR4, con aplicación en estudios de diseño de compuestos orgánicos dirigidos a esta diana farmacológica.

3) Integración de los protocolos de modelado por homología y de simulación de MD desarrollados en las etapas anteriores, en un flujo de trabajo automático ofrecido como servicio web abierto a la comunidad científica.

Recientemente, se han desarrollado diversos servidores web que ofrecen la predicción de la estructura de proteínas problema, algunos de ellos adaptados al caso de las proteínas de membrana en general, y otros incluso a los GPCRs en particular. En nuestro laboratorio hemos integrado las metodologías para modelado y simulación de GPCRs descritas anteriormente en el único flujo de trabajo automatizado en el servidor GPCR-ModSim (<http://gpcr.usc.es>). Esto por un lado aumenta la competitividad de nuestro grupo de investigación, además de ofrecer dicho servicio de forma libre a la comunidad científica, por lo que usuarios carentes de un extenso bagaje en el modelado molecular pueden obtener fiables modelos por homología de GPCR. En la publicación de dicho

RESUMEN

servidor, se realizó una comparación con otros servicios análogos, obteniendo excelentes resultados en cuanto a la predicción de la estructura del receptor de histamina H₁ justo después de su publicación, además de mostrar el máximo nivel de interacción con el usuario en las distintas etapas para producir modelos por homología (incluyendo un alineamiento de secuencia automatizado); aportando como novedad la posibilidad de realizar simulaciones de MD de GPCRs. El servidor está en constante actualización, tratando de ofrecer soluciones computacionales a la caracterización estructural y dinámica de GPCRs, con aplicaciones en el ámbito bioquímico y farmacológico de estos receptores.

Como conclusión, en la presente tesis se han empleado técnicas del estado-del-arte en el modelado molecular y en el descubrimiento y diseño de fármacos asistido por ordenador en diversos proyectos, con la caracterización estructural de GPCRs como hilo conductor. Los resultados obtenidos cuentan tanto con exitosas combinaciones metodológicas como su aplicación en el estudio de la bioquímica, química médica y farmacología de miembros la superfamilia de los GPCRs.