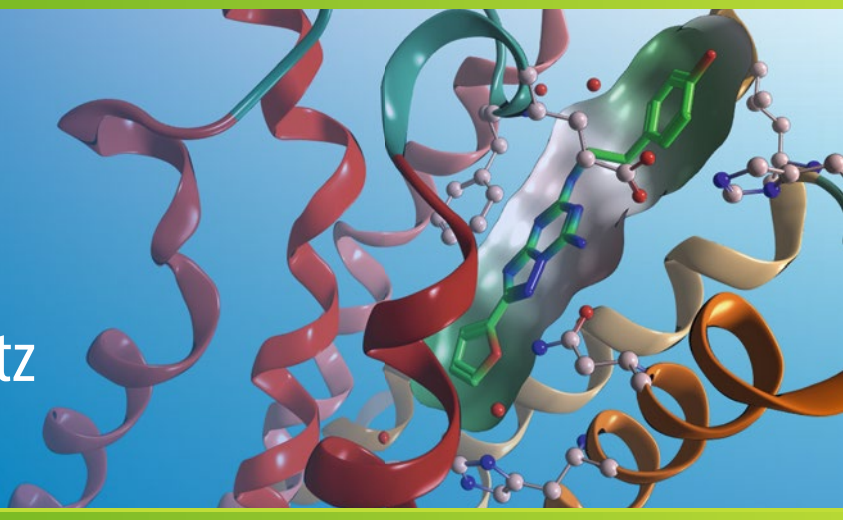
Methods in Molecular Biology 1705

Springer Protocols



Alexander Heifetz
Editor

Computational Methods for GPCR Drug Discovery

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Computational Methods for GPCR Drug Discovery

Edited by

Alexander Heifetz

Evotec (UK) Ltd., Abingdon, Oxfordshire, UK

Humana Press

Editor
Alexander Heifetz
Evotec (UK) Ltd.
Abingdon, Oxfordshire, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-7464-1 ISBN 978-1-4939-7465-8 (eBook)
https://doi.org/10.1007/978-1-4939-7465-8

Library of Congress Control Number: 2017957977

Printed on acid-free paper

This Humana Press imprint is published by Springer Nature
The registered company is Springer Science+Business Media, LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

G protein-coupled receptors (GPCRs) have enormous physiological and biomedical importance, being the primary site of action of approximately 40% of prescribed drugs. Although the human genome encodes more than 850 different GPCR proteins, to date drugs have only been developed against 50 of these. Thus, there is a unique opportunity to design new therapies for a huge number of unexploited but potentially tractable targets. Recent advances in GPCR pharmacology and structural biology together with developments in computational modeling have resulted in a resurgence in the number of GPCR drug discovery campaigns.

This book provides a unique overview of modern computational strategies and techniques employed in the field of GPCR drug discovery, including structure- and ligand-based approaches and cheminformatics. It is demonstrated how these computational approaches can be used to address key issues in drug discovery such as receptor structure modeling, protein-ligand interactions, GPCR function, flexibility and dynamics, ligand binding kinetics, positions of water molecules and their role in ligand binding, calculation of the free energy of binding (affinity), prediction of the effects of mutations on ligand binding, interconversion between agonists and antagonists, deorphanization of GPCRs, and discovery of biased and allosteric modulators. A review of these techniques will allow a diverse audience, including structural and molecular biologists, computational and medicinal chemists, pharmacologists and drug designers, to navigate through and effectively deploy these advances.

Abingdon, Oxfordshire, UK *Alexander Heifetz*

Contents

Contributors

Matteo Aldeghi • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*

Patricia Angelloz-Nicoud • *Research Informatics, Evotec (France) SAS, Toulouse, France*

Johan Åqvist • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*

Damian Bartuzi • *Department of Synthesis and Chemical Technology of Pharmaceutical Substances with Computer Modelling Lab, Medical University of Lublin, Lublin, Poland*

Philip C. Biggin • *Department of Biochemistry, Structural Bioinformatics and Computational Biochemistry, University of Oxford, Oxford, UK*

Mike J. Bodkin • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*

Andrea Bortolato • *Heptares Therapeutics Ltd., Welwyn Garden City, Hertfordshire, UK*

Ewa I. Chudyk • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*

Antonella Ciancetta • *Molecular Recognition Section, Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA*

Jason B. Cross • *University of Texas MD Anderson Cancer Center, Houston, TX, USA*

Constantino Diaz • *Research Informatics, Evotec (France) SAS, Toulouse, France*

Iwan J.P. de Esch • *Department of Medicinal Chemistry, Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

Mauricio Esguerra • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*

Dmitri G. Fedorov • *CD-FMat, National Institute of Advanced Industrial Science and Technology (AIST), Ibaraki, Japan*

Sławomir Filipek • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*

Marta Filizola • *Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA*

David Gloriam • *Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark*

Chris de Graaf • *Department of Medicinal Chemistry, Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*

Dong Guo • *Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, Xuzhou Medical University, Xuzhou, Jiangsu, China*

Hugo Gutiérrez-de-Terán • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*

Paul C.D. Hawkins • *OpenEye Scientific Software, Santa Fe, NM, USA*

George Hedger • *Department of Biochemistry, Structural Bioinformatics and Computational Biochemistry, University of Oxford, Oxford, UK*

Alexander Heifetz • *Evotec (UK) Ltd., Oxfordshire, UK; Division of Biosciences, Research Department of Structural and Molecular Biology, Institute of Structural and Molecular Biology, University College London, London, UK*

Adriaan P. IJzerman • *Division of Medicinal Chemistry, Leiden Academic Centre for Drug Research (LACDR), Leiden University, Leiden, The Netherlands*

Kenneth A. Jacobson • *Molecular Recognition Section, Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA*
Jakub Jakowiecki • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
Tim James • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*
Willem Jespers • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*
Agnieszka A. Kaczor • *Department of Synthesis and Chemical Technology of Pharmaceutical Substances with Computer Modelling Lab, Faculty of Pharmacy with Division of Medical Analytics, Medical University of Lublin, Lublin, Poland; Department of Pharmaceutical Chemistry, School of Pharmacy, University of Eastern Finland, Kuopio, Finland*
Albert J. Kooistra • *Department of Medicinal Chemistry, Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; Centre for Molecular and Biomolecular Informatics (CMBI), Radboud University Medical Center, Nijmegen, The Netherlands*
Paul Labute • *Chemical Computing Group Inc., Montreal, QC, Canada*
Dorota Latek • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
Kristen A. Marino • *Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
Jonathan S. Mason • *Heptares Therapeutics Ltd., Welwyn Garden City, Hertfordshire, UK*
Dariusz Matosiuk • *Department of Synthesis and Chemical Technology of Pharmaceutical Substances with Computer Modelling Lab, Medical University of Lublin, Lublin, Poland*
Przemysław Miszta • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
Inaki Morao • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*
Ana Oliveira • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*
Daniel F. Ortwine • *Discovery Chemistry, Genentech, Inc., South San Francisco, CA, USA*
Emilie Pihan • *Research Informatics, Evotec (France) SAS, Toulouse, France*
Andrew Potterton • *Structural and Molecular Biology, University College London, London, UK; Evotec (UK) Ltd., Oxfordshire, UK*
Roger Robinson • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*
Ismael Rodríguez-Espigares • *Department of Experimental and Health Sciences, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Pompeu Fabra University (UPF), Barcelona, Spain*
Ewelina Rutkowska • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
Jessica Sallander • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*
Laurie Sarrat • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*
Jana Selent • *Department of Experimental and Health Sciences, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Pompeu Fabra University (UPF), Barcelona, Spain*
Robert T. Smith • *Heptares Therapeutics Ltd., Welwyn Garden City, Hertfordshire, UK*
Michelle Southey • *Evotec (UK) Ltd., Abingdon, Oxfordshire, UK*

Gunther Stahl • *OpenEye Scientific Software, Köln, Germany*
Tomasz Maciej Stepniewski • *Department of Experimental and Health Sciences, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Pompeu Fabra University (UPF), Barcelona, Spain*
Christofer S. Tautermann • *Department for Medicinal Chemistry, Boehringer Ingelheim Pharma, GmbH & Co KG, Biberach an der Riss, Germany*
Benjamin G. Tehan • *Heptares Therapeutics Ltd., Welwyn Garden City, Hertfordshire, UK*
Sid Topiol • *3D-2drug, LLC, Fair Lawn, NJ, USA; Center for Healthcare Innovation, Stevens Institute of Technology, Hoboken, NJ, USA*
Andrea Townsend-Nicholson • *Division of Biosciences, Research Department of Structural and Molecular Biology, University College London, London, UK*
Maria Turant • *Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Warsaw, Poland*
Silvana Vasile • *Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden*
Márton Vass • *Department of Medicinal Chemistry, Amsterdam Institute for Molecules Medicines and Systems, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
Naushad Velgy • *Department of Biochemistry, Structural Bioinformatics and Computational Biochemistry, University of Oxford, Oxford, UK*
Stefan Verhoeven • *Netherlands eScience Center, Amsterdam, The Netherlands*

Chapter 1

Current and Future Challenges in GPCR Drug Discovery

Sid Topiol

Abstract

GPCRs play a pervasive physiological role and, in turn, are the leading target class for pharmaceuticals. Beginning with the determination of the structure of rhodopsin, and dramatically accelerating since the reporting of the first ligand-mediated GPCR X-ray structures, our understanding of the structural and functional characteristics of these proteins has grown dramatically. Deploying this now rapidly emerging information for drug discovery has already been extensively demonstrated through a watershed of studies appearing in numerous scientific reports. Included in these expositions are areas such as sites and characteristics of ligand to GPCR binding, protein activation, effector bias, allosteric mechanisms, dimerization, polypharmacology and others. Computational chemistry studies are demonstrating an increasing role in capitalizing on the structural studies to further advance our understanding of these proteins as well as to drive drug discovery. Such drug discovery activities range from the design of orthosteric site inhibitors through, for example, allosteric modulators, biased ligands, partial agonists and bitopic ligands. Herein, these topics are outlined through specific examples in the hopes of providing a glimpse of the state of the field.

Key words GPCR, Structure-based drug discovery, X-ray structure, Allosteric modulators, Receptor bias

1 Introduction

As early as 50 years ago, when the first computer software programs were being written, computational chemistry tools were being developed to understand and guide drug properties and discovery. Ligands for G-protein coupled receptors [GPCRs], such as endogenous amines, and psychotropic drugs such as LSD acting on these receptors, were a common focus of these research efforts. Electrostatic point charge representations of ligands were often used, followed by molecular mechanics, low level quantum chemical (e.g., semi-empirical), and later ab initio methods. Calculations of static properties such as atomic point charges, electrostatic fields, and electron densities of small molecules pushed the envelope of computational hardware and software wherein LSD was a "large" molecule and the structure of its GPCR target receptors, such as

Alexander Heifetz (ed.), *Computational Methods for GPCR Drug Discovery*, Methods in Molecular Biology, vol. 1705, https://doi.org/10.1007/978-1-4939-7465-8_1,

the 5-HT receptors, could only be imagined. Computational studies involving the target GPCR proteins were beyond reach both because of the computational limitations presented by such large systems and the lack of useful structural information about these proteins. For decades following this, computational chemistry software and the available hardware capabilities grew dramatically allowing for far more sophisticated, accurate, and rapidly generated information. On the experimental side, significant advances in areas such as molecular biology, protein crystallography and structure determination, and NMR provided the means for detailed investigations of proteins at the atomic level. These experimental approaches have been successfully applied to soluble proteins for many years so that both small molecule and protein computational methods were deployed and advanced extensively for these targets. For membrane bound proteins such as GPCRs, the powerful approach of X-ray structure determination remained elusive, thus limiting computational drug discovery to ligand-based methods such as pharmacophore studies. While ligand-based methods have indeed been very successful, the much sought atomic level structural information, with its more powerful and far-reaching capabilities, remained a much sought after goal. The first glimpse of the architecture of these proteins came from electron microscopy studies of the related 7 transmembrane protein, bacterio-rhodopsin [1, 2]. While efforts were made to use this structure as a template for homology models of GPCRs of interest, the distal relationship between them did not allow for the suitable accuracy of models needed for drug discovery. It was not until 2000 that the first X-ray structure of a GPCR, the class A GPCR rhodopsin, was reported [3]. Although this was not a ligand-mediated GPCR, it provided a significant advance in the information needed for understanding the structure and function of GPCRs, especially for class A GPCRs. Extensive use was made of the structure of the transmembrane region of rhodopsin as a template for homology models for ligand-mediated GPCRs, but the greater structural accuracy needed for the most efficient drug design was still not achieved. This was exacerbated by the far more varied structure of the extracellular loops of these proteins which contributed to the differential involvement of this extracellular loop region which generally interacted directly with bound ligands at the orthosteric sites. The biggest informational breakthrough for ligand-mediated GPCRs came with the X-ray structure determinations of the first ligand-mediated (class A) GPCR, those of the β2-adrenergic receptor (β_2AR) [4, 5]. These first detailed atomic level structure reports of a ligand-mediated GPCR heralded the beginning of a new era of computer-aided drug discovery for GPCRs. The details of the orthosteric ligand's binding, including the involvement of residues from the extracellular loops, were seen in these first examples as well as in various functional features of GPCRs such as the so-called

"ionic lock" and "tryptophan switch" which had previously been characterized and/or hypothesized by other experimental studies. Computational methods that had matured over years as applied to mostly soluble proteins were directly and instantly deployed toward these targets. The determination of these GPCR X-ray structures, and others solved since then, has involved a number of methods to overcome the challenges of crystallizing membrane bound proteins. Prominent among these for facilitating crystallization were the use of companion proteins which were either covalently bound (e.g., T4L, BRIL) at either side of the 7-TM (e.g., spliced into IC3 or attached at N-terminus) or non-covalently bound (e.g., antibodies and nanobodies), selective stabilizing mutations, bound high-affinity ligands, and the lipidic cubic phase methods. The power of structure-based drug-discovery (SBDD) when applied to these first structures was quickly demonstrated. In the now 10 years since those structures were reported, there have been numerous other X-ray structures reported covering examples of a number of class A subclasses, various activation states, other classes of GPCRs (B, C, and F) and yielding a watershed of tools for drug discovery and understanding of the detailed molecular mechanisms and parameters governing a host of physiological roles. Among the many class A X-ray structures, that of LSD bound to 5-HT_{2b} is now added to the arsenal [6] and begins to satisfy the imagination that has stirred over many years.

2 GPCR Structure: A Bottoms-Up Guided Tour

The term GPCRs refers to a broad range of proteins with a common architectural feature, i.e., a domain consisting of seven alpha helices which traverse the cell membrane alternatively from the extracellular (EC) side to the intracellular (IC) side (helix 1 or H1 or TM1) and back again (helix 2) etc. Helices 1 and 2 are connected on their intracellular side by an intracellular loop (IC1) while helices 2 and 3 are connected on their extracellular side by extracellular loop 2 (EC2) with corresponding connecting loops and nomenclature for all the helices. This description generally defines class A GPCRs, whereas non-class A GPCRs (classes B, C, and F) contain an additional extracellular domain. The most well-established effector proteins to which these proteins couple are G-proteins, which is the source of the name "GPCR." As it is now well established that an important role of these proteins is to couple to other effectors besides G-proteins a more universal name seems called for. Often, the name "7TM" is used for all of these proteins, owing to their common architectural feature. Nevertheless, as the name GPCR remains widely used and recognized for all of these proteins, we will use it herein.

As the primary role of GPCRs is by definition to couple with G-proteins, or more generally with effectors, on the intracellular side of the membrane bound GPCRs, we begin our tour of the architecture of GPCRs there. Various regions are highlighted which have gained recognition for their structural, functional, or ligand binding roles.

2.1 The Intracellular Rim

With the role of 7TM proteins to induce signal propagation to the intracellular region via interaction with their various effectors, this region serves as the initial conduit for this information transmission mechanism. The structures of the 7TM proteins and their changes in this region determine whether fruitful interactions with the effectors will take place (activation), to what extent these effective interactions will occur (intrinsic activity), and with which effectors these will occur (biased agonism). The various X-ray structures now available, together with a wealth of molecular biological, biophysical, and biochemical studies, include examples spanning these various possibilities. At the fully active protein extreme is the X-ray structure of the fully activated β_2AR receptor [7] in complex with a high-affinity agonist BI-167107 and its effector, the heterotrimeric GTP binding protein Gs. As with many of the 7TM X-ray structures, companion proteins used to aid in the crystallization are included in the structure. Here, there are two such proteins, the camelid nanobody Nb35 and T4L (replacing the N-terminus of the 7TM). In this case, the role of the camelid nanobody in helping to stabilize the active form of the 7TM protein was demonstrated through molecular dynamics simulations [8] an approach playing an increasing role in complementing X-ray structural information. In comparison with structures of the inactive state, this structure reveals a more extended conformation for helix 5 and an outward shift of helix 6 from the central helical transmembrane axis while helices 3 and 7 move slightly inward. Similar structural information for the transmembrane region is available for the fully inactive protein extreme which has generally been more accessible due to the greater availability of high affinity antagonists (versus agonists) to facilitate protein crystallization as illustrated by an X-ray structure of the adenosine 2a receptor (A_{2a}AR) [9]. Rhodopsin X-ray structures of the inactive state provide other, earlier examples. In addition to the X-ray structures of the active and inactive extremes, there are now a number of examples of various intermediate states including complexes with partial agonists, and demonstrating intermediate structural features. It is noteworthy that the first X-ray structures of GPCR proteins were those of rhodopsin, due in large part to the availability of large quantities of the protein for crystallization. Thus, while ligand-mediated GPCR (non-rhodopsin) proteins hold a central focus for much of the interest in this area because of their pharmacological role as drug targets, rhodopsin has played an early and

continued role in unraveling structure/function information about 7TM proteins. A recent example of this is the reporting of an X-ray structure of rhodopsin bound to the β-arrestin effector [10] which starts to shed light on the GPCR structural differences corresponding to differential effector interaction, i.e., biased signaling. Compared to the β_2AR /Gs active structure, H6 in the Rhodopsin/arrestin active structure exhibits a 4**Å** lesser outward shift (10**Å** vs. 14**Å**). Supporting the relevance of this difference not being an artifact of the difference in GPCRs (rhodopsin versus β_2AR) smaller differences are also seen in TM1, TM4, TM5, and TM7 when comparing the rhodopsin/arrestin active structure to the active state structure of rhodopsin bound to a C-terminal peptide of Gα.

2.2 The Most Intracellular Ligand Site—To Date

While the role of these proteins is to communicate information from the extracellular region, generally via ligand (or light, in the case of rhodopsin) mediated signaling, to the intracellular region via interaction with various effectors, the location for the signal modulating ligand has traditionally been understood to be in the upper region of the protein for orthosteric as well as sites bordering these orthosteric sites (acting as selectivity sources or allosteric sites). In striking contrast to this, chemokine receptor X-ray structures for CCR2 and CCR9 demonstrate that inhibitor binding in the extreme IC region of the protein and immediately proximal to the effector binding region occurs [11, 12]. The CCR9 X-ray structure has only one bound ligand, the inhibitor vercirnon, which is bound at this site and juts out at the IC domain. Two simultaneous inhibitor ligands are bound in the CCR2 X-ray structure. The first inhibitor, CCR2-RA-[R], is bound in the same location as vercirnon in the CCR9 X-ray structure (*see* Fig. 1), whereas the second ligand, BMS-681, is bound in the assumed orthosteric site. The inhibitory role of the ligands at this extreme IC location is reflected in the protein structure wherein the outward movement of H6, required for the effector binding to the 7TM protein, is prevented by the inhibitor. Additionally, the inhibitor's position directly precludes effector binding. This IC region allosteric site uncovered in these studies is not known to have any endogenous role, and can be considered an illustration of a *man-made* site [15].

2.3 The B Site

The X-ray crystal structure of corticotropin-releasing factor receptor 1 (CRF_1R), a member of the secretin like class B GPCRs, in complex with the antagonist CP-376395 [16] revealed yet another *man-made* 7TM ligand binding site (B site) which is also much deeper than the classical orthosteric site. The location of CP-376395, a compound identified through screening studies, is further away from the IC region of the protein toward the EC

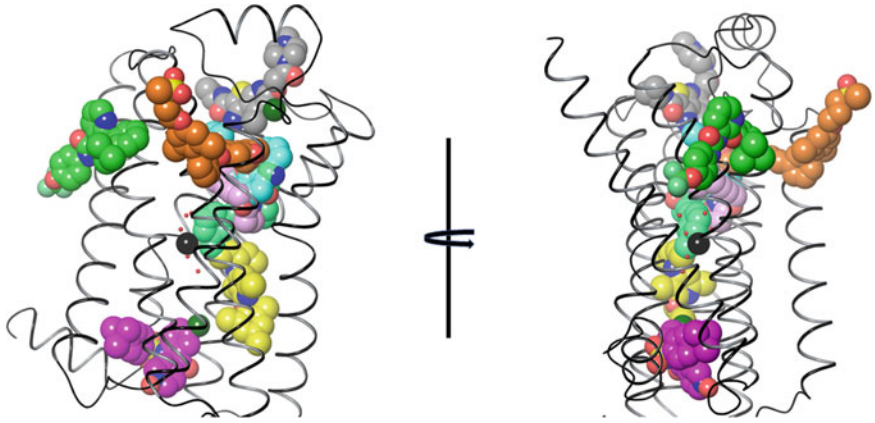


Fig. 1 Illustration of the range of GPCR ligand binding sites. Examples of binding sites of selected ligands in the 7TM domain, as seen from the side of the α-helical barrel. The ligands are superimposed with a ribbon representation of the 7TM domain using the X-ray structure of β_2AR. The identity of each ligand, the protein to which it is bound, and the Protein Data Bank (PDB) [13] accession number of the complex are as follows: vercirnon (purple carbon atoms) in CCR9 (PDB:5LWE); CP-376395 (*yellow* carbon atoms) in CRF_1R (PDB:4K5Y); sodium/water cluster (*black* sodium atom, *red* water molecules) in the A_{2a}AR (PDB:4EIY); mavoglurant (*spring green* carbon atoms) in mGluR5 (PDB:4OO9); iperoxo (*plum* carbon atoms) in the M2 receptor (PDB:4MQT); carazolol (*aqua* carbon atoms) in the β_2AR (PDB:2RH1); LY2119620 (*gray* carbon atoms) in the M2 receptor (PDB:4MQT). Two views are shown at 90° rotation as indicated. The molecular graphics were generated in Maestro [14]

region than that of the ligands in the CCR2 and CCR9 X-ray structures described above (*see* Fig. 1). Nevertheless, it is still far removed from the orthosteric binding sites expected for this class of proteins wherein peptide-like ligands are expected to bind in, e.g., an open, v-shaped EC cavity as found in the X-ray structure of the related class B receptor whose X-ray structure has also been solved [17]. CP-376395 is selective for CRF_1R over CRF_2R. Two characteristics of this site would suggest conflicting predictions regarding its potential as a source of selectivity. Pervasive dogma argues that allosteric sites offer greater opportunities for selectivity than orthosteric sites, a principle based on the expected conservation of residues among related proteins for common ligands at their orthosteric sites. In contrast, there is generally expected to be less variation in structure and sequence of 7TM proteins in the IC direction than the EC direction. In this case, differences in just two residues at this binding site between CRF_1R and CRF_2R could provide the explanation for the greater preference observed with CP-376395 for CRF_1R. Analysis of this structure also suggests that CP-376395 prevents the activating outward motion of TM6, thereby explaining its inhibitory effect and offering clues for design of ligands with desired intrinsic activity [16, 17].

2.4 The Ionic Lock ("D(E)RY")

As with a much of the early understanding of 7TM structure/function relationships, evidence for this feature as a characterization of the inactive state has its origins in rhodopsin X-ray structures. Using the Ballesteros-Weinstein numbering scheme [18], the ionic lock describes the structure of the cluster of residues D3.49, R3.50, Y3.51 ("DRY"), and E6.30 as a means for establishing the inactive state of the protein. In the structure of the inactive state of rhodopsin, R3.50 interacts with D3.49, E6.30, and T6.34. The ionic interaction of R3.50 with E6.30 forms the lock between helices 3 and 6 which is associated with the inactive state. X-ray structures of the inactive state of ligand-mediated class A GPCRs, e.g., the D3 dopamine receptor [19] have been found to include this ionic lock. Interestingly, similar ionic and/or polar hydrogen bonding networks are found in X-ray structures of inactive forms of class B [16, 17], class C [20, 21], and class F [22, 23] GPCRs. In active state structures of rhodopsin and the β_2AR [7, 24] the interaction of R3.50 with E6.30 is no longer present, but R3.50 interacts with Y5.58 instead. A number of X-ray structures with common ligands but varying ionic cluster interactions, along with X-ray structures with ligands of varying intrinsic activity, and molecular dynamics simulations, lead to an emerging picture that these active/inactive state ionic lock indicators are not guarantees of the activation state but serve as indicators of their propensities for the given state [15]. Moreover, they seem to contribute to the induction of the structural changes more proximal to the effector.

2.5 Internal Water Network and Its Sodium Site

Sodium has been shown to act as an allosteric modulator of 7TM proteins. A 1.8-Å high-resolution X-ray structure of the A_{2a}AR with the inhibitor ZM241385 bound [25] shows the position of a sodium atom at the center of a network of water molecules which traverse much of the transmembrane region and has three clusters whose central cluster contains the sodium atom (Fig. 1). This cluster is situated between the ionic lock and the so-called toggle switch (*see* below). This site can potentially serve as a ligand binding site as supported by a crystal structure of the 7TM region of a class C GPCR, the mGluR5 receptor [21] containing the bound negative allosteric modulator (NAM) mavoglurant whose lower portion overlaps spatially with this sodium/water cluster (Fig. 1). In the case of the A_{2a}AR, the orthosteric site is located in the more common upper region of the transmembrane as is the inhibitor also seen in the A_{2a}AR X-ray structure. For mGluR5 however, the orthosteric site resides in an extracellular domain separated from the 7TM domain by a "cysteine-rich" protein linker. The mavoglurant site in mGluR5 is thus another example of a *man-made* site [26]. The role of this sodium/water-cluster region to serve as an allosteric site to two very differently located orthosteric sites is more uniformly understood when viewed as serving a common function to modulate the same local transmembrane region.

In the case of the A_{2a}AR, active state structures are available for comparison [27, 28] and show that the hydrated sodium-ion induces a kinking in helices VI and VII. Ameloride is known to compete with this site for the A_{2a}AR and it has been used for structure-based design [29–33].

2.6 The CWxP "Toggle Switch"

Analogous to the ionic lock, a highly conserved CWxP motif contains tryptophan W6.48 whose orientation had been hypothesized as a marker for the activation state of 7TM proteins. This has now been verified extensively in numerous GPCR X-ray structures where there is a shift in the position of the indole of W6.48 between the active and inactive state structures, albeit not a flipping of the indole ring as originally hypothesized. Interestingly, the driving forces for this indole positioning are varied. In the inactive state structures of rhodopsin [3], the histamine H1 receptor [34], and the muscarinic M2 receptors [35], the inhibitors (retinal in the case of rhodopsin) hold the corresponding indole of W6.48 in the same position by directly interacting with it. In other instances, such as the inhibitor bound inactive state X-ray structures of the β_2AR [4] or the dopamine D3 receptor [19], ligand interaction is with an aromatic ring of an intervening residue. Whereas this region is proximal to the endogenous ligand's binding sites in class A 7TM proteins, that is not the case for class C 7TM proteins such as mGluRs. It is thus interesting that X-ray structures of the 7TM domains of mGluRs show that allosteric inhibitor bound proteins with ligands at this man-made site (for mGluRs) [20, 21] have their corresponding tryptophan rings displace outward from the 7TM core by the bound ligand. The role of the differences in the structural features in this region in protein activation is becoming clearer. Comparing the active and inactive states of the A_{2a}AR [27, 36] shows that the agonist sits much deeper in the pocket forming a series of hydrogen bonds with the protein as well as a steric clash with W6.48 which collectively induce more active like orientations and positions of H5 and H6.

2.7 The "Orthosteric" Pocket—The HUB

The approximately upper third region of the 7TM core generally serves as the binding site for endogenous ligands, particularly for class A 7TM proteins and, in turn, for most synthetic ligands; herein referring to this as the "HUB" region. In considering all GPCRs, many more ligands are accommodated at the HUB than effectors at the IC region. It is therefore intuitive that there is considerable diversity at this HUB site as described above. This diversity is rooted in multiple sources including significant amino acid variability between 7TM proteins in the EC direction, still greater variability in the 7TM connecting extracellular loops and greater structural variation (such as degree of openness) in this region.

The architecture of this HUB region makes it the region of first choice for ligands in general and for endogenous ligands for similar reasons. In addition to its relative diversity, it is for the most part an enclosed cavity, yet not as narrow as the more internal section of the 7TM which is therefore more limiting in the scope of ligands it can accommodate. It is also generally accessible from the extracellular side—the source of most modulators.

Relative to the 7TM helical axis, ligands are found to span a considerable vertical depth and lateral breadth of this HUB region, which may reasonably be considered as comprised of sub-regions. In some cases, such as inhibitors of the histamine H1 receptor [34] and the muscarinic M2 [35] and M3 [37] receptors, ligands reside at the extreme depth of this region (*see* Fig. 1), having no direct interaction with the extracellular loops that are often a source of selectivity as well as affinity. Nevertheless, single-residue differences in this region can account for selectivity of inhibitors between closely related proteins such as M2 and M3 [15]. More generally, ligands binding in the HUB region interact with the extracellular loops as well (*see*, e.g., the many structures for the β_2AR or the A_{2a}AR) and interactions with portions of the N-terminal regions are also seen (*see*, e.g., Ref. 38). In the lateral direction, small molecules can bind in nonoverlapping positions at the same depth within the 7TM domain (*see* Fig. 2). For example, the inhibitor AZD1283 in a P2Y12 X-ray structure [39] where AZD1283 laterally spans the TM region interacting with helices III-VII, and the inhibitor in the CB1 X-ray structure [38] occupy nonoverlapping locations as shown in Fig. 2.

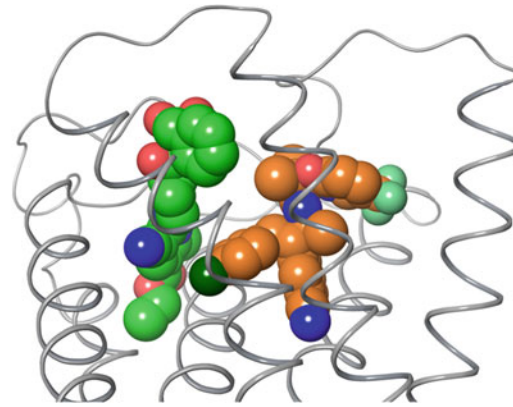


Fig. 2 Illustration of laterally parallel sub-sites at the orthosteric HUB site. Locations of two ligands which bind in the HUB site of the 7TM domain, as seen from the side of the α-helical barrel. The ligands are superimposed with a ribbon representation of the 7TM domain using the X-ray structure of β_2AR. The identity of each ligand, the protein to which it is bound, and the Protein Data Bank (PDB) [13] accession number of the complex are as follows: taranabant (*orange* carbon atoms) in CB1 (PDB:5 U09); AZD1283 (*green* carbon atoms) in the P2Y12 receptor (PDB:4NTJ). The molecular graphics were generated in Maestro [14]

While the structural characteristics of the orthosteric site are clearly related to the affinities of ligands at these sites, the more subtle sources of the varying intrinsic activities of these ligands are becoming clearer. Based on reported protein X-ray structures in different states for the same proteins, e.g., the adenosine and adrenergic GPCR proteins, the more inactive states are associated with more open pockets in this region containing slightly larger ligands, with increasing activating influence and decreasing openness of the site as one proceeds from inverse agonists through partial agonists to full agonists [40]. The relative openness here plays a role in structural changes propagating to the IC region related to the relevant active/inactive state. This trend is not universal as the ultimate criteria lie in the induced changes in the IC direction. Yet more subtle than understanding of intrinsic activity is the understanding of the mechanism of ligand bias for ligands binding here and inducing different effector interaction profiles. The structural bases for underlying biased agonism are emerging. For example, a comparison of the X-ray structure complexes of 5-HT_{1b} and 5-HT_{2b} [41, 42], each bound with ergotamine, a compound having bias for β-arrestin over G-protein at 5-HT_{1b} shows differences at the binding site in conformations of residues at helix 6 which, in turn, correspond to a less active like conformation of helix 6 in the IC region. The crystal structure of LSD bound to 5-HT_{2b} [6] adds to this picture and, coupled with molecular dynamics studies, indicates interaction with extracellular loop 2 (ECL2) modulates the LSD off rate thereby providing a kinetic component to enhanced β-arrestin interaction.

2.8 The EC Rim (Vestibule, Address Site, Etc.)

The upper rim of the 7TM domain has the most diverse features that are reflected in the variability in the types of ligands residing there as well as the nature of its usage. Ligands binding here range from small molecules to relatively large peptides. Small ligands are found occupying this region as seen, e.g., for the A_{2a}AR and *CXCR4* receptors. This region serves as an allosteric "vestibule" as in the case of muscarinic receptors [35, 37, 43] (Fig. 1). Regions above or below this region can combine with it to form binding sites for ligands. Together with the region below it, it is thus utilized as an "address" pocket in conjunction with ligands binding their "message" portions more deeply into the 7TM such as has been seen for structures of the opioid receptors [44–47]. Alternatively, ligands bound here can be found to also interact with portions of the protein in the N-terminus direction for class A (e.g., CB1 [38]) and, e.g., class B proteins where structural and mutational evidence indicates that endogenous ligand binding straddles the 7TM and EC domains [17].

2.9 Binding To and Through the External 7TM Wall

The overall barrel-like shape of the transmembrane region of the 7TM proteins, whose most commonly accepted orthosteric binding site lies inside the barrel-like structure, is consistent with a simple model for the trajectory of ligands engaging in interactions with these proteins. An alternate model for this trajectory has been considered for some time, wherein a ligand approaches and enters a 7TM protein from the outer side of the barrel [48–50]. With examples of X-ray structures showing ligands completely buried within the 7TM and covered by EC loops now available for, e.g., rhodopsin [3], S1p1 [51], and PAR1 [52], an external trajectory becomes a more plausible explanation for ligand entry. Indeed, there is now proof of ligands actually binding partially external to, or even completely external to the 7TM region (*see* GPR40 [53] and $P2Y_1$ [54] respectively). External interactions as modulators of GPCR activity are supported by other types of data. Both homo- and hetero-dimerization is known to play a role in the functioning of various GPCRs [55–58]. Binding of cholesterol to the external transmembrane region has been shown by X-ray structures (*see*, e.g., [4, 59]) as well as electron microscopy [60] and is believed to play a role in dimerization. Long time frame molecular dynamics investigations are helpful in examining these potential interactions [61] and new mass spectrometry-based tools are emerging to measure the dependence and degree of protein oligomerization due to membrane lipid binding mediation [62]. As GPCR signaling is dependent on changes in their helical positions and conformation in the IC regions to prepare for effector interaction, it is not surprising to find evidence that modulating such changes from the external side of their transmembrane region is possible. Taken together, the collection of structural information that is now available for 7TM proteins indicates that *all regions of 7TM proteins, inside and out, appear to provide potential sites for ligand modulation*.

2.10 EC Domains for GPCR Classes B, C, and F

The class B, C, and F 7TM receptor subgroups are differentiated in part from class A 7TM receptors by an additional domain at their N-terminus, the EC domain. X-ray structures for the 7TM domains have been determined for examples of all three of these protein subgroups and confirm their generally similar architecture to the class A GPCR proteins. While peptidic ligands for the class B 7TM proteins bind in between the 7TM and EC domains, the EC domains of the class C and F 7TM proteins contain structurally separated binding sites. Class C receptors contain a cysteine-rich linker region that connects the 7TM domain to the so-called Venus "flytrap" (VFT) domain to which the endogenous ligands bind, such as glutamate in the case of mGluR receptors. X-ray structures of the EC domains of class C and class F receptors with ligands bound have been determined. For the EC domain, the structural changes associated with the active versus inactive state of the VFT,

as modulated by ligands, are understood. While X-ray structures of the VFT domains of mGluRs were first reported in 2000 [63], they have not been used extensively for structure-based design, albeit SBDD studies including additional X-ray structures are still being reported [64]. The belief that the transmembrane region offers an advantageous environment for drug discovery is responsible for this, although the rationales given, such as greater sequence variation in the 7TM domains, have been challenged [65]. There is no full-length structure of a Class C GPCR protein available at this time to establish the interplay between these domains. For the class F 7TM proteins, full-length structures have now been determined for the example of the smoothened receptor [66]. The smoothened proteins have a cysteine-rich C terminal domain which is connected through a linker to its 7TM domain. Structural insights into ligand-mediated communication from the 7TM to the EC domains have emerged and the stage is set for understanding the EC to 7TM communication mechanism as well.

3 Computational Drug-Discovery Approaches/Capabilities

The exploitation of X-ray structures of ligand-mediated 7TM proteins through computer-aided drug discovery followed rapidly and with striking successes after these structures became available and have been reviewed extensively. Three major factors have contributed to the precipitation and rapid growth of 7TM SBDD work. First, experimental tools and strategies, such as the use of companion proteins, lipidic cubic phase methods, and targeted stabilizing mutations, enabled the crystallization and X-ray structure determination of 7TM (and other classes of membrane bound) proteins. Second, computational chemistry and modeling methods, notably for protein structure studies, had been honed over many years with the available protein targets. While these targets were primarily soluble proteins, many of the computational approaches were independent of the differences. Third, the inherent architecture and most commonly targeted HUB ligand binding site are highly suitable for small molecule binding and modulation. The enclosed nature, probably not coincidentally, is essentially ideal for occupancy by a small molecule and in striking contrast to other extremes such as surface binding sites, inter-domain sites, etc. The "druggability" of this HUB site, the broad range of endogenous ligand modulators utilizing this HUB, the prominence of GPCR targeted drugs, and the pervasive role of GPCR proteins can all be argued to stem from these GPCR architectural characteristics.

3.1 Virtual Screening, High-Throughput Docking

As was demonstrated soon after the first X-ray structures for GPCR proteins were reported, high-throughput docking (HTD) campaigns using X-ray structure models for compounds acting at those targets, and in the same fashion (site, active/inactive state, etc.) are extremely effective for GPCR proteins. From a drug discovery perspective, there is little doubt that this is an extremely efficient and striking approach to begin studies of a target protein. Screening of large databases of preexisting compounds, from commercial or proprietary sources, can rapidly jump start a drug discovery program with identification of potent compounds and structure activity information. In silico HTD screening of large databases containing millions of compounds, using a number of different software systems, is already routinely conducted and used to rank and select as few as tens or hundreds of compounds for in vitro testing. Hit rates above 30% and yielding compounds with activities in the single-digit nanomolar range are common [15]. Integration of HTD methods with ligand-based methods or protein-based pharmacophore methods often further improves these successes. As expected, the success of these approaches depends on how close to this optimal paradigm one operates. Within a subgroup of closely related targets for proteins with common endogenous ligands (e.g., adrenergic or dopaminergic receptors) homology models based on X-ray structures of other members of the subgroup yield comparable results to those where the X-ray structure of the target of interest is used. However, in silico screening for an agonist using an antagonist bound structure of the same protein as a template is often more challenging than for an antagonist using a homology model based on an X-ray structure template of another protein in an inactive state within the same subgroup. This is because differing states of a protein have greater deviation in their binding site structures from their templates than common states within a subgroup where there are few amino acid changes. As one progresses to create and deploy homology models based on templates of X-ray structures outside a target sub-group the reliability of the homology model decreases. In part, this is due to the reduction in sequence identity, and consequentially reduction in structural similarity in the transmembrane helices. More elaborate protocols such as the use of multiple templates help improve the accuracy of the homology models. More significantly, the extracellular loops, and in some instances sections of the N-terminal, vary much more significantly in shape, length, fold etc., while contributing significantly to ligand binding at the HUB.

3.2 Structure-Based Drug Design

The strengths and weaknesses described for HTD pertain more generally to computational drug discovery for GPCR proteins. It has become commonplace to employ models of 7TM proteins in drug discovery activities as evidenced by the extensive reporting of these approaches in the medicinal chemistry literature. Unlike

HTD studies where the sheer number of ligand structures being investigated necessitates invoking considerable restrictions and approximations on the methodologies employed, studies of individual complexes allow for more sophisticated approaches to be employed leading to more accurate results. Thus, while most HTD studies use frozen, or nearly frozen protein structures in calculations with highly approximate models for the interaction energetics, this is not the case otherwise. Considerable protein flexibility ranging from the active site to the entire protein complex structure is common using molecular mechanics based approaches. More accurate quantum chemical ab initio methods are possible for more accurate evaluations of, e.g., ligand protein interactions, albeit this has not been as extensively reported. To explore processes such as conformational changes associated with activation, molecular mechanics-based molecular dynamics methods for extended time frame simulations are being reported more regularly. Additionally, complementing the increasing structural information on the involvement of explicit water molecules in various regions of GPCRs such as orthosteric sites or the sodium binding site seen in A_{2a}AR crystal structures [25], computational methods are now more reliably examining the various roles of these waters. Finally, the increased capabilities of GPCR modeling are opening up new types of opportunities such as the use of homology modeling of orphan receptors to identify ligands for use in de-orphanizing these receptors [67].

4 Scope of Information: Lessons and Emerging Opportunities

The evermore rapidly emerging structural and functional information for GPCRs, especially from X-ray structure determinations, translates into various properties and processes of these systems of relevance to drug action and is now becoming more amenable to investigation and exploitation by computational approaches. The various aspects of drug discovery dominate the general interests here, with protein structure/function insights sought as well and these two goals are often inseparable. It is interesting to consider the range in scope, and potential already realized as well as anticipated.

4.1 Allosteric Sites: Another Look at the Other Site

The use of X-ray structures of GPCRs for discovery and design of ligands in the simplest approach, i.e., for the *same site and same activity* as the X-ray structure, is now well established and remarkably effective. As noted, at the most common (for class A) 7TM orthosteric HUB site, structural differences observed between inactive and active state structures provide a clear explanation for the deterioration of results when inactive structures are used (without other moderations or considerations) to identify activating compounds. It is reasonable to assume that similar differences

would occur at other ligand binding sites, i.e., allosteric sites, such as those described herein, but there is as yet not much data available to test this. The HUB site serving as the orthosteric site for class A proteins, however, serves as an *allosteric* site for non-class A proteins. X-ray structures for the smoothened class F GPCR, with activators or inhibitors bound reveal differences between their binding sites [22, 23]. For the mGluRs, while X-ray structures in the 7TM domain are only available with inhibitors bound, SAR data finds that very small ligand changes result in the switch between activating and inactivating ligands [21, 26, 68]. Whether this apparent sensitivity is inherent in the role of this site as an allosteric site or simply a consequence of the still limited information is unclear. Relatedly, whereas by definition an allosteric site is a site other than that where the endogenous ligand binds (the orthosteric site), other implications for the allosteric nomenclature, e.g., the modulation of the orthosteric binding site events, may point to a different perspective. GPCRs are a category of proteins whose architecture and prominent functioning can be described as proteins whose communication with intracellular effectors is generally modulated by ligand interaction at the control-center/HUB site. It therefore seems logical to re-consider the EC domain sites of non-class A GPCRs as *operationally allosteric* sites as compared to their more unifying (with respect to class A GPCRs) HUB. By analogy to class A GPCRs, these 7TM sites would be directly involved in signal transmission in the IC direction as opposed to the usual indirect model wherein these 7TM sites modulate signaling in the EC direction (at the EC domain) which must then propagate back through the same 7TM domain—where they began. Evidence for both perspectives exists vis. the X-ray structure of the complete smoothened protein [66] shows evidence for ligand binding in the 7TM domain resulting in interactions from the 7TM with the EC domain which influence EC ligand binding whereas evidence for the direct ligand control at the IC region of non-class A GPCR proteins is provided by reports of a truncated mGluR5 protein without its EC domain which can be activated by a TM binding ligand [69]. The inherent machinery of GPCRs thus questions whether the usual roles, experimental analyses, and ligand design of allosteric modulators should be treated differently for non-class A GPCRs.

4.2 Multi-Target Tuning: Within and Between Subgroups, to Other Classes, Tuning In Vs. Out, Polypharmacology

A critical factor in the action of drugs is the profile of their activity at varying targets. This target profile is important even when only simple inhibition is considered at multiple targets and extends to considerations of varying activities at different sites of different proteins such as activators with respect to one site and inhibitors with respect to another. Indeed, poly-pharmacology has grown as a medicinal approach. The growing structural information that has become available now introduces a broad selection of opportunities