



Universitat Autònoma de Barcelona

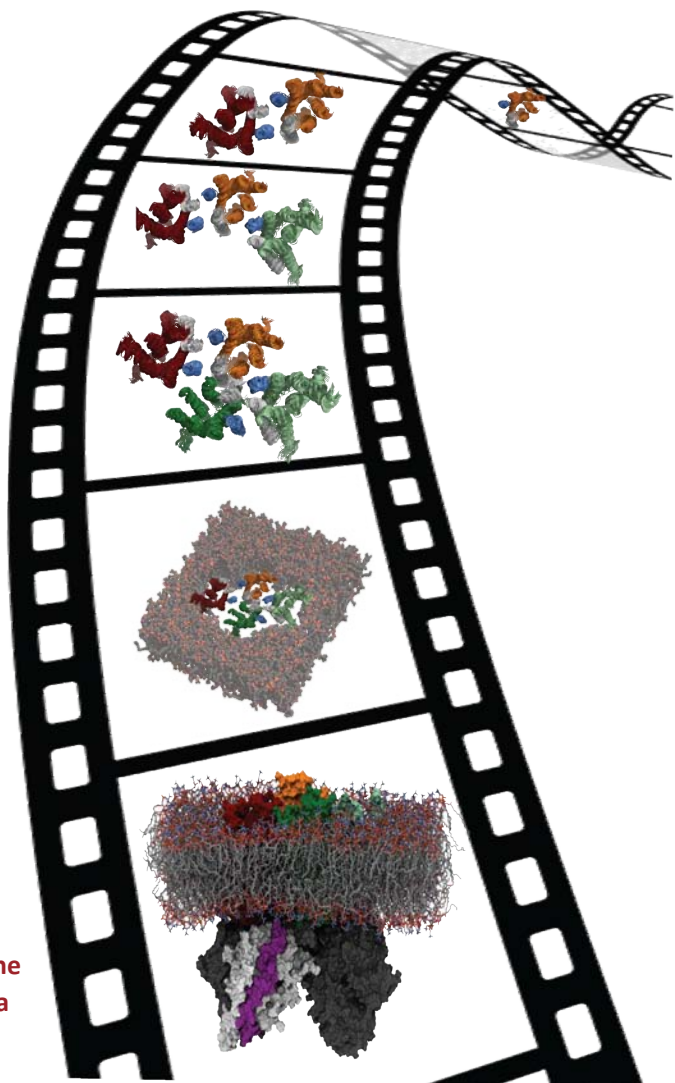
ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

Application of Molecular Dynamics methods to the study of biological systems

Laura Perez Benito
2016



PhD Thesis

Laboratory of Computational Medicine
Universidad Aut3noma de Barcelona



Universitat Autònoma de Barcelona

Medicine Faculty
Bioestatics Department

Application of Molecular Dynamics methods to the study of biological systems

Laura Pérez Benito
November 2016

PhD in Biochemistry, Molecular Biology and Biomedicine

Directors:

Leonardo Pardo Carrasco, Arnau Cordoní Montoya, Gary Tresadern

Tutor: Dr. Mireia Duñach



Universitat Autònoma de Barcelona

Medicine Faculty

Bioestadística Department

Application of Molecular Dynamics methods to the study of biological systems

This work has been carried out by Laura Pérez Benito under the supervision of Prof. Leonardo Pardo Carrasco at Autonomous University of Barcelona, Arnau Corderó Montonya Postdoctoral Researcher at Autonomous University of Barcelona and Gary Tresadern Senior Principal Scientist, Discovery Sciences Janssen Research & Development, to obtain the Degree of Doctor in Biochemistry, Molecular Biology and Biomedicine.

Dr. Leonardo Pardo Carrasco

Dr. Arnau Corderó Montonya

Dr. Gary Tresadern

Acknowledgements

The group has changed a lot since I first entered. On my first day I met Jose Carlos. I took my old Macbook with me and I remember his first sentence: “so the group is divided into two groups: Linux users and Mac users, I guess I know which one you are gonna chose”. I started in the other office sharing with Santi (sitting in Arnau’s table at that moment) Minos, Julian and Marc. I also remember their first sentence: “Google is your friend and if you ask any question make sure it is a smart one”. It can sound a bit rude but it was just the beginning and of course it has been totally different from then on.

First of all I would like to thank Leo. He was the one that gave me the opportunity to do the PhD. I contacted Leo, who I met through my colleagues in Almirall, and I had an interview with him. It was a funny interview and it ended up with a comment: “Well the guys in the group are asking me to have a girl because right now all the group is just guys.” In the beginning this also sounds a bit bad, but then getting to know Leo you realize this is just part of his humor.

Leo is a good director, a good boss, a good chef (the best tuna I have ever tried!) and most of all a good person to have a chat that can go from

GPCRs to more personal problems and concerns. I have learnt a lot from him and every time I go into his office I still do. Thank you for all.

I want to thank Arnau. He is the most patient person ever, and the person from which I have learnt all the methodology needed to perform this thesis. He always has time for a quick chat in the office to solve doubts or during the car rides in the LMC car!

I would also like to thank Gary. My supervisor during my stay in Janssen and also co-director of this thesis. Thanks for poking me and helping me to finish stuff, to make me think that there is always a more perfect way of doing things, to look deeper into the computational problems not just the superficial, and of course helping me in moments of stress.

I would like to thank Julian, Minos and Marc, office colleagues. Thanks for helping me to get comfortable with Modeller and Gromacs and also with IT problems while Arnau was away, and of course for all the beers and tequila shots shared!

I also want to thank Edu, office colleague in the last three years. I always enjoy our little jokes in the office, youtube videos like teletubies and blancanieves! Wait for the small teletubie dance after the thesis ☺.

Also thank him the last few months solving problems with the computers, the water scripts... He is always available to solve and help!

I'd like also to thank Santi. Santi, the lunches without your comments are not the same! While the telocogresca is always fun with you!

I'd like to thank Jose Carlos. You have always found time to solve problems or talk about science in general and have always been really helpful with all kind of scripting issues. I can't forget of course our common hobby: talking about the Simpsons and quoting every time needed! Thanks also for all the Apolo nights and for your mum's Paella, the best ever!

Thanks also to Angel and Gianlu (both located in the baby room office), thanks for the coffee breaks, for listening to my complaints and for always having good advice and posing difficult questions during all the evaluations, they have been challenging but really useful.

Quiero dar las gracias a mis padres, por apoyarme y por enseñarme a luchar por lo que quiero. Y a mis hermanas por haber sido un referente y un ejemplo para mí a pesar de estar lejos, por hacer crecer en mí las ganas de leer y aprender.

Before finishing, I would like to thank one more person. Thanks to Alberto, for introducing me to the field of computational chemistry, for supporting me during my degree, and for all the chats over these years.

‘To start press Any Key’. Where’s the ANY key?

Homer J. Simpson

Table of contents

1.	<u>Introduction to Molecular Dynamics</u>	1
	1.1. Applications	4
	1.2. Free-energy perturbations	6
	1.3. References	10
	1.4. Objectives	13
2.	<u>Methods</u>	15
	2.1. Homology modeling	16
	2.2. Protein and ligand preparation	20
	2.3. Docking	22
	2.4. MD simulations	27
	2.4.1. Preparation and execution of a MD simulation	32
	2.5. Free-energy calculations	33
	2.6. References	38
3.	<u>Application of Molecular Dynamics Simulations</u>	45
	3.1. Introduction to G protein-coupled receptors	46
	3.2. The structure of G protein-coupled receptors	48
	3.3. GPCR activation	50
	3.4. Allosteric modulators	53
	3.5. Oligomerization of GPCRs	56
	3.6. mGlu ₂ receptor as drug target	61

3.7. References	65
3.8. Quaternary structure of a G protein-coupled receptor heterotetramer coupled to Gi and Gs	74
3.8.1 Background	74
3.8.2. Results and discussion	77
3.8.3 Conclusions	94
3.8.4. Methods	96
3.8.5 Supporting information	104
3.8.6. References	120
3.9. The transmission switch mechanism of allosteric modulation of the metabotropic glutamate 2 receptor	124
3.9.1. Background	124
3.9.2. Results	129
3.9.3. Discussion	147
3.9.4. Conclusions	158
3.9.5. Methods	159
3.9.6. Supporting information	170
3.9.7. References	176
4. <u>Application of Free-energy perturbations</u>	<u>185</u>
4.1. Alzheimer's disease	186
4.2. AD hypotheses	187
4.3. A β production and β -secretase	188
4.4. BACE1 as a drug discovery target	190
4.5. BACE1 inhibitors	194
4.6. References	201

4.7. The Application of Free energy perturbation for the design of BACE1 inhibitors	206
4.7.1. Background	206
4.7.2. Results	212
4.7.3. Discussion and conclusions	227
4.7.4. Experimental section	242
4.7.5. Supporting information	251
4.7.6. References	261
5. <u>Conclusions</u>	<u>269</u>
5.1. The quaternary structure of the adenosine A ₁ -A _{2A} receptor heteromer	270
5.2. Allosteric modulators of class C GPCRs	271
5.3. Design selective beta-secretase-1 inhibitors	271
6. <u>List of publications</u>	<u>273</u>

1. Introduction

One of the principal computational tools in the theoretical study of biological molecules is the method of molecular dynamics simulations (MD). This computational approach permits the study of the evolution of a system in time, giving us insights on processes that occur in biological systems, at atomic scales, such as the role of protein flexibility in ligand binding.

The MD method was first introduced by Alder and Wainwright in the late 1950's^{1,2} and performs integration of Newton's second law of motion to calculate the forces at sequential time steps and thereby displacing atoms accordingly. They studied the interactions of hard sphere motions during a 9.2 ps total simulation. Several systems were calculated using MD but it was not until 1977 when the first protein simulations appeared, investigating a folded globular protein, the bovine pancreatic trypsin inhibitor, with total simulation times of 8.8 ps.³ Since then, in parallel with the increasing impact of computational methods in many areas of science, vast improvements have been seen. Improved theory, methodology and hardware now mean that since the late 1990's MD is increasingly used to understand chemistry and biology of protein systems. Furthermore, in the last decade the impressive fast technological development in computational speed and data storage volume

permits simulations to be performed reaching the time scales of many biologically relevant processes,⁴ (Figure 1). For instance, it is common to see modern reports of simulations with micro- to millisecond timescales enabling side chain rotations and loop motions to be investigated.^{5,6} But, making longer simulations is not enough to extend the conformational sampling in biomolecular systems. Most of the simulations just explore a small region around the energy minimum closest to the initial conformation. Taking advantage again of the technological advances, an obvious strategy is to perform a series of parallel simulations with several starting conformations.⁷ Mathematical approaches are used to analyze the amount of trajectories generated, clustering the results.^{8,9}

Hence, the field has evolved from studying isolated macromolecules in vacuum in a picosecond time scale to studying complex biomolecular systems composed of millions of atoms with simulation time scales spanning up to milliseconds. Nowadays computational chemistry is still undergoing significant changes due to access and porting of algorithms to Graphics Processing Unit (GPU) hardware which contain thousands of cores in a cheap highly parallel architecture that is efficient for computational approaches such as MD.^{10,11}

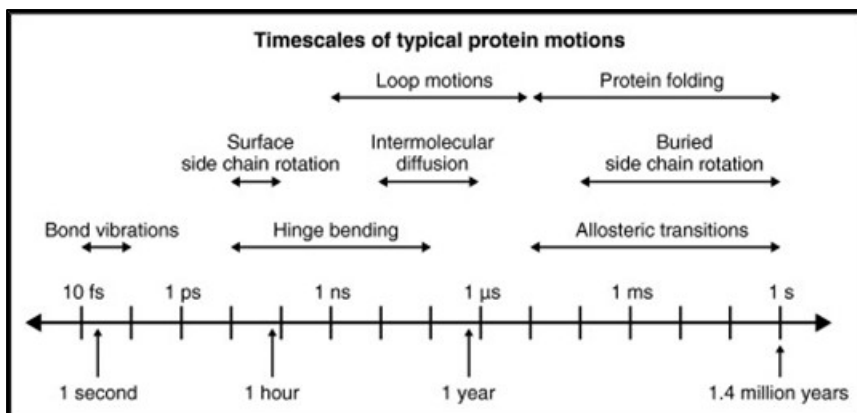


Figure 1. Timescales for typical protein conformational changes. Below the axis is a rough estimate of the amount of ‘wallclock’ time required to perform a molecular dynamics (MD) simulation of a typically sized protein–protein complex solvated in explicit water (45,000 atoms) on a typical (2.6 GHz dual-core) desktop computer. Taken from Zwier and Chong Curr Opin. Pharmacol, 2010, 10, 745-752.

1.1. Applications

The use of MD simulations is now an essential technique enabling more relevant events to be observed *in silico* for the rationalization of experimental data. One of the most common uses of MD has focused on studying ligand-macromolecule complexes. For instance, modeling in reverse the activation process of a membrane protein such a G-protein coupled receptor (GPCR), starting from an active state and simulating towards the inactive conformation and thereby witnessing the atomistic changes required for large scale

protein motion associated with G-protein binding.¹² Further, the same β_2 -adrenergic receptor system has been studied with other MD methods, such as the Markov State Models to investigate the differences between agonist and antagonist interaction with the activation pathways.¹³

The application to drug-target complexes, can be used in conjunction with X-ray crystallographic structures, or homology models to elucidate the binding mode of a particular ligand at a given protein target.¹⁴ The resulting trajectories are analyzed to extract information such as distances and interactions between atoms or residues of interest, and can be used to generate hypotheses for subsequent molecular design, examples include studies of muscarinic receptors,¹⁵ and sphingosine-1-phosphate receptor 1.¹⁶ MD can also overcome the major limitations of static structure-based drug design and in particular the limitation of rigid docking calculations, which do not sample the protein conformational changes observed during ligand binding.¹⁶ During unbiased MD simulations, the model system evolves freely over time and the binding site can be sampled with different amino acid side chain and backbone conformations that may be required for binding of some molecules. Therefore MD simulations are an ideal

way to obtain multiple conformations of macromolecular targets that can then be used, for instance, for ensemble docking.¹⁷ Also, the dynamic nature of proteins is now increasingly understood¹⁸ and working only with single X-ray crystallographic snapshots can hinder understanding of the mechanism of action of drug molecules. In this sense MD also provides an advantage to analyze and interpret conformational effects which are likely important, particularly for molecules which may have allosteric functional effects.¹⁹

1.2. Free energy perturbation

Knowledge of the energetics of ligand binding would facilitate successful completion of drug discovery projects. In this regard, state of the art methods such as docking have proven to perform poorly in this task as they were typically designed to reproduce binding modes or enrichment for virtual screening, but not predict the relative affinities of analogues from a lead chemical series. Building on the advances in MD has brought Free-Energy perturbation (FEP) calculations closer to reality for application in drug discovery projects. Hence, free-energy simulations are a

valuable approach. This includes methods such as FEP that employ molecular dynamics or Monte Carlo (MD/MC) simulations to assess the free-energy difference between two related ligands via either a chemical or alchemical path.^{20,21} FEP has previously shown to be a promising *in silico* technique to estimate binding affinities.²² In fact, in a typical lead optimization program, the calculation of the relative difference in binding energy between two compounds is a key parameter. Interestingly, this relative difference is more easily computed than the ‘absolute’ binding free-energy of a single compound due to inherent difficulties in accurately computing components of the thermodynamic cycle such as absolute solvation energies. Instead, the FEP method performs the alchemical perturbation between two close analogue molecules in solution and in protein (see Figure 2), processes A and B are difficult to simulate using molecular dynamics (Figure 2 panel left), so to determine $\Delta(\Delta G)$ computationally, an alternative route is used (Figure 2 panel right). These new paths mutate the ligand A into ligand B in the unbound and bound states and hence many expected errors cancel due to the similarity of the two systems. FEP calculations *per se* are not new. Based on the ideas of Zwanzig,²³ FEP was applied in the 1970s and 1980s, when a number of research groups presented the

first concepts of free-energy methods (e.g., McCammon *et al* in Nature and Bash and co-workers in Science).^{3,24} However, lack of computational power and limited parameterization in early force fields, impeded substantial progress in the field despite its attractiveness.²⁵

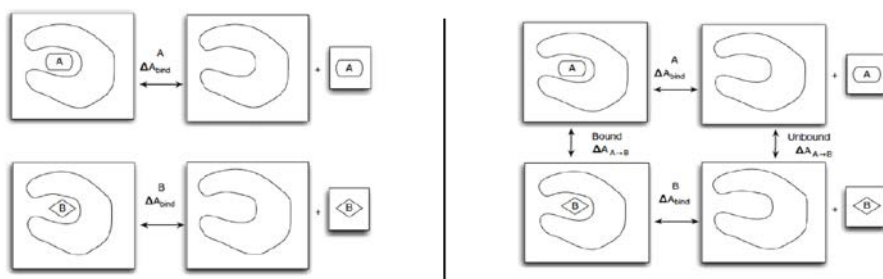


Figure 2. The FEP thermodynamic cycles showing the perturbation performed between two analogous molecules in solution and in protein. Adapted from Biomolecular Simulations Methods and Protocols.²⁶

Recent advances such as better force fields, novel sampling algorithms, and low-cost (GPU-based) compute power can now deliver the level of accuracy and speed required for a typical drug discovery project.^{22,27,28} Combined with the large increase in crystallographic structures available in the public domain (Protein Data Bank)²⁹ and inside many pharmaceutical companies this opens the door for routine application of structure-based methods such as

FEP, in drug discovery. Hence the pharmaceutical industry is highly interested in applying this approach and many avenues are needed to be explored to fully understand its use for fast moving lead optimisation projects.

1.3. References

- 1 Alder, B. J. & Wainwright, T. E. Phase Transition for a Hard Sphere System. *The Journal of Chemical Physics* **27**, 1208-1209, (1957).
- 2 Alder, B. J. & Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *The Journal of Chemical Physics* **31**, 459-466, (1959).
- 3 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).
- 4 Mortier, J. *et al.* The impact of molecular dynamics on drug design: applications for the characterization of ligand-macromolecule complexes. *Drug Discovery Today* **20**, 686-702, (2015).
- 5 Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond time scale using molecular dynamics simulations. *The Journal of General Physiology* **135**, 555 (2010).
- 6 Pierce, L. C. T., Salomon-Ferrer, R., Augusto F. de Oliveira, C., McCammon, J. A. & Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *Journal of Chemical Theory and Computation* **8**, 2997-3002, (2012).
- 7 Ferruz, N., Tresadern, G., Pineda-Lucena, A. & De Fabritiis, G. Multibody cofactor and substrate molecular recognition in the myo-inositol monophosphatase enzyme. *Scientific Reports* **6**, 30275, (2016).
- 8 Bowman, G. R., Huang, X. & Pande, V. S. Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* **49**, 197-201, (2009).
- 9 Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology* **18**, 154-162, (2008).
- 10 Harvey, M. J., Giupponi, G. & Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *Journal of Chemical Theory and Computation* **5**, 1632-1639, (2009).

- 11 Páll, S. & Hess, B. A flexible algorithm for calculating pair interactions on S-IMD architectures. *Computer Physics Communications* **184**, 2641-2650, (2013).
- 12 Dror, R. O. *et al.* Activation mechanism of the β 2-adrenergic receptor. *Proceedings of the National Academy of Sciences* **108**, 18684-18689 (2011).
- 13 Kohlhoff, K. J. *et al.* Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat Chem* **6**, 15-21, (2014).
- 14 Clark, A. J. *et al.* Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *Journal of Chemical Theory and Computation* **12**, 2990-2998, (2016).
- 15 Miao, Y. *et al.* Accelerated structure-based design of chemically diverse allosteric modulators of a muscarinic G protein-coupled receptor. *Proceedings of the National Academy of Sciences* **113**, E5675-E5684 (2016).
- 16 Stanley, N., Pardo, L. & Fabritiis, G. D. The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Scientific Reports* **6**, 22639, (2016).
- 17 Campbell, A. J., Lamb, M. L. & Joseph-McCarthy, D. Ensemble-Based Docking Using Biased Molecular Dynamics. *Journal of Chemical Information and Modeling* **54**, 2127-2138, (2014).
- 18 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007).
- 19 Hertig, S., Latorraca, N. R. & Dror, R. O. Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLoS Comput Biol* **12**, e1004746, (2016).
- 20 Homeyer, N., Stoll, F., Hillisch, A. & Gohlke, H. Binding Free Energy Calculations for Lead Optimization: Assessment of Their Accuracy in an Industrial Drug Design Context. *Journal of Chemical Theory and Computation* **10**, 3331-3344, (2014).
- 21 Hansen, N. & van Gunsteren, W. F. Practical aspects of free-energy calculations: a review. *Journal of Chemical Theory and Computation* **10**, 2632-2647 (2014).
- 22 Wang, L. *et al.* Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and

- Force Field. *Journal of the American Chemical Society* **137**, 2695-2703, (2015).
- 23 Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *The Journal of Chemical Physics* **22**, 1420-1426 (1954).
- 24 Bash, P., Singh, U., Langridge, R. & Kollman, P. Free energy calculations by computer simulation. *Science* **236**, 564-568 (1987).
- 25 Chodera, J. D. *et al.* Alchemical free energy methods for drug discovery: progress and challenges. *Current opinion in structural biology* **21**, 150-160 (2011).
- 26 Michael R. Shirts, D. L. M. *Biomolecular Simulations Methods and Protocols*. 271 -311 (Springer New York Heidelberg Dordrecht London, 2013).
- 27 Anderson, J. A., Lorenz, C. D. & Travesset, A. General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics* **227**, 5342-5359 (2008).
- 28 Harder, E. *et al.* OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation*, (2015).
- 29 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).

1.4. Objectives

The aim of this thesis is to apply molecular dynamics (MD) simulations to answer contemporary biochemical challenges. This includes the use of computer-aided drug discovery techniques to develop therapeutically important molecules.

The thesis is organized in two sections.

The first section focuses on using all-atom MD simulations to understand the structure and function of G protein-coupled receptors (GPCRs). Thus, the first aim of this thesis is understanding how GPCR (hetero) oligomerization influences the physiological role of receptors, and their use as therapeutic targets (chapter 1). Despite the proven success of GPCRs as drug targets, useful ligands do not exist for the majority of them. The main reason being that the orthosteric binding sites across members of a GPCR subfamily for a particular endogenous ligand are often highly conserved, making it difficult to achieve high selectivity for specific GPCR subtypes. Novel approaches to modulate GPCRs, to overcome this problem, involve the discovery of multivalent ligands that target physiologically relevant GPCR hetero-oligomers (chapter 2), or allosteric ligands that bind at allosteric sites and act in conjunction with the endogenous ligand (chapter 3). Thus, the

knowledge of the structural elements involved in GPCR (hetero) oligomerization achieved in chapter 1 is translated into predictive tools for the selection and design of molecules targeting physiologically relevant GPCR heteromers (chapter 2). Recent publications of crystal structures of metabotropic glutamate receptors (class C GPCRs) led us study how negative (NAMs) and positive (PAMs) allosteric modulators bind and, in particular, how the ligand exerts its allosteric functional effects (chapter 3).

The second section focuses on the use of Free Energy Perturbations (FEP) methods for the design of beta-secretase 1 (BACE1) inhibitors.

2. Methods

The term Molecular Modeling is used to describe the study of molecules using physical models. The group of rules used to describe a system is denominated theory, and is often expressed with mathematical equations. The aim of combining a theoretical formalism with a physical model is double: it allows rationalizing experimental data and also provides a tool that is able to predict the behavior of a system when lacking experimental information.

The following sections describe the methods used in this thesis. The topics covered include: Homology modeling; Docking; Protein and ligand preparation; MD simulations and FEP calculations. The specific details for the methods employed in each of the studies will be described within each chapter.

2.1. Homology modeling

Homology modeling is the process of constructing a 3D atomic model of a target protein normally based on the known structure of a similar protein. Perhaps the largest number of applications has been in the field of membrane proteins such as ion channels and GPCRs, due to the difficulties to solve crystallographic structures for these protein families, despite

their huge pharmaceutical interest. In the following paragraphs we describe some specific considerations for homology modeling of GPCRs and their complexes. It is relevant for this thesis given our interest to study a Class A GPCR such as the A1R-A2AR heterotetramer for which structure is unavailable, but also Class C mGlu receptors which will be amongst the more difficult systems we study with this technique.

The number of detailed 3D protein structures deposited in the Protein Data Bank (PDB)¹ continues to increase in a significant manner (Figure 1).

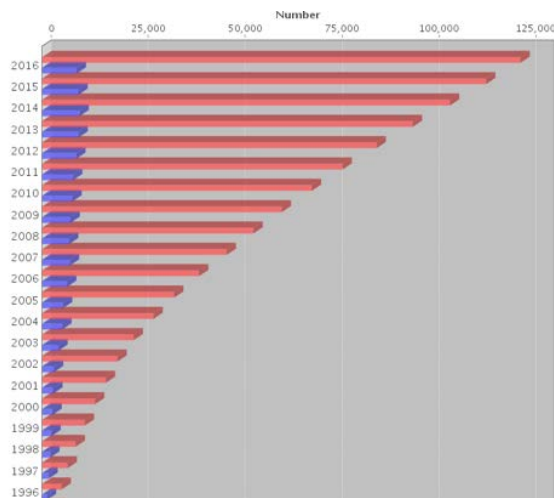


Figure 1. Yearly growth of total structures. Extracted from PDB.¹

In the field of GPCRs, innovative crystallographic techniques have increased the number of GPCR structures resulting in a golden age for structure-based drug design on this class of membrane proteins.² Structures are available for Class A, B, C, and F receptors, in complex with agonists, inverse agonists, antagonists, or allosteric modulators, bound to intracellular proteins such as the G-protein³ (also G protein fragments and G-protein mimicking nanobodies)⁴ or arrestin.⁵ Despite this continuous increase of solved GPCR structures, for the vast majority of pharmaceutically relevant receptors, structural information is accessible only by cost-efficient alternatives like homology modeling.⁶

Development of homology models is a multi-step process (Figure 2), that can be summarized in the following way (1) identification of a template; (2) perform single or multiple sequence alignments; (3) model building for the target based on the 3D structure of the template; (4) model refinement, analysis of alignments (including gap deletions and additions) and (5) model validation.

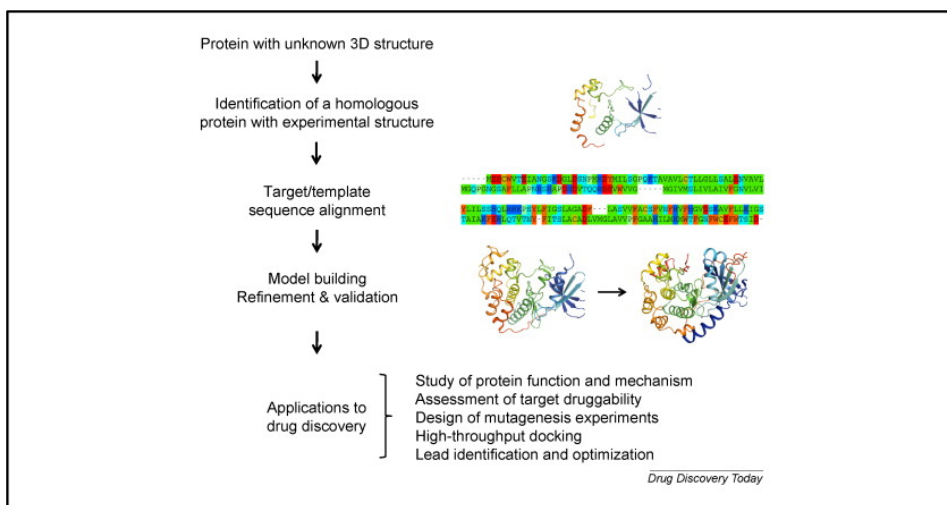


Figure 2. Summary of the multi-step process for homology modeling. Taken from Cavasotto and Phatak *Drug Discovery Today*, 14, 2009, 676-683 .7

One such tool which is suitable for building homology models is the program MODELLER,⁸ which has been mostly used in this thesis. The program uses homologous structures to construct constraints on atomic distances, dihedral angles, and so forth, these are then combined with statistical distributions derived from many pairs in the PDB. The software combines the sequences and structures into a complete alignment which can then be examined using molecular graphics programs and edited manually. Alternatively, similar methods are available (and have been used in this thesis) in the MOE software suite.⁹ The homology modeling tools in MOE share some similarities with MODELLER and the

overall workflow is to **1.** Specify the initial target geometry coming from one or multiple template structures **2.** Handle insertions and deletions in the sequence **3.** Loops and sidechain modeling and packing **4.** Final model selection and refinement is performed based on the considerations such as the RMSD of the intermediate models, the solvation energy, knowledge-based assessment of the packing quality, contact energies. Both MODELLER and MOE permit building homology models in the presence of bound ligands. This is a useful feature to maintain a binding site which can then be used for molecular docking.

2.2. Protein and ligand preparation

Before using crystallographic or homology-modeled structures in simulations it is important to pre-process both protein and ligand. This step, which is sometimes overlooked, is crucial for each accurate computational results.^{10,11} With regards to protein preparation, issues to consider include filling gaps or loops that were unresolved in crystal structures, ionization of amino acids, correct rotamer conformations, and correct hydrogen placements which is particularly important for binding site water molecules.

Correspondingly, ligand preparation presents similar concerns with regards to ionization, tautomerization and correct parameterization for the corresponding molecular mechanics force fields. Hence, the methods used for protein and ligand preparation are dependent on the approaches that are to be used for subsequent work, for instance the software and force field used in MD and FEP calculations.

When performing MD calculations we have typically applied the following approach. Ligand atom types are assigned with the appropriate force field in mind, and the partial charges have been obtained using HF/6-31G*-derived RESP¹² atomic charges in Gaussian.¹³

Although Maestro¹⁴ tools have been used for preparing protein for the subsequent Glide docking,¹⁵ in the case of MD simulations the proteins were prepared using AMBER,¹⁶ to add Hydrogens, missing side chains and cap the N- and C-termini.

For FEP calculations we also used Schrodinger Maestro software, hence the protein and ligands were prepared in the following way that is similar to the protocol used for docking. For the protein, structure preparation was performed using the Protein Preparation

Wizard with default settings to fix missing sidechains/atoms, assign protein protonation states with PROPKA,¹⁷ optimize the hydrogen bonding network, assign ligand charges, and relax crystal contacts with a brief minimization to RMSD 0.5 Å. The catalytic aspartates present in the active site were treated in their ionized states. Ligands were ionized and then docked into the binding site. Docked poses were used as input for FEP calculations, but manual inspection was always performed, and in some cases the dihedral angles of terminal aromatic rings need to be corrected to ensure they are consistent with the entire set of input ligands. In short, some manual curation of input to FEP is always recommended.

2.3. Docking

Computational docking is the process of placing a ligand in a binding site in a plausible binding mode and assessing the quality (or energetics) of the fit. Computational pose prediction with docking is possible with different strategies.¹⁸⁻²⁰ However, assessing the quality of the fit is regarded highly challenging, because the scoring functions employed are highly simplified empirical approaches to assess binding energies.²¹⁻²³ Ultimately, docking

scoring functions are deliberately parameterized to be simple in order to keep them as fast as possible to enable large scale virtual screening of millions of compounds.

Over the last three decades structure-based methods have taken a prime place in the drug discovery process.^{24,25} Docking is typically used for virtual screening to computationally differentiate actives from inactives to find hit molecules for a target protein. In a simple way this approach depends on successfully discriminating the molecules which can fit, or bind, at the site of interest versus those with little chance. Docking is also used in drug discovery lead optimization of analogues but the analysis of results is typically a subjective task because the scoring functions are not deemed accurate enough to rely upon. Hence, while useful to differentiate between actives and decoys, docking has been relatively unsuccessful in the prediction of binding affinity.^{26,27} Recent developments in terms of docking algorithms have primarily focused on more accurate inclusion of (de)solvation effects. Examples of these improved docking algorithms include WScore and WaterFLAP (Molecular Discovery).²⁸ Improving the

(de)solvation effects have indeed increased the ability of docking to better distinguish between (highly) actives and inactives.

In this thesis two approaches have been used for docking: MOE and GLIDE docking. In the former, the ligand placement method is based on the Triangle matcher algorithm. Poses are generated by aligning ligand triplets of atoms on triplets of alpha spheres in a more systematic way than in the Alpha Triangle method. The binding poses are scored based on the London dG approach.²⁹ The energy functional is a sum of the following terms:

$$\Delta G = c + E_{flex} + \sum_{h-bonds} c_{HB} f_{HB} + \sum_{m-lig} c_M f_M + \sum_{atoms\ i} \Delta D_i \quad (1)$$

Where c represents the average gain/loss of rotational and translational entropy; E_{flex} is the energy due to the loss of flexibility of the ligand (calculated from ligand topology only); f_{HB} measures geometric imperfections of hydrogen bonds and takes a value in $[0,1]$; c_{HB} is the energy of an ideal hydrogen bond; f_M measures geometric imperfection of metal ligations and takes a value in $[0,1]$; c_M is the energy of an ideal metal ligation; and D_i is the desolvation energy of atom i .

For GLIDE docking the Maestro software was used. GLIDE relies upon the pre-calculation of a fixed grid that narrows the search space then the next step is to perform a torsionally flexible energy optimization on an OPLS-AA³⁰ non-bonded potential grid with subsequent refinement via a Monte Carlo pose sampling method. Ligands are placed with an algorithm based on Emodel,³¹ to select between protein-ligand complexes of a given ligand scoring function and GlideScore³² function, to rank-order compounds to separate those that bind strongly (actives) from those that don't (inactives). The GLIDE XPEXtra-Precision Glide,¹⁵ scoring function was used in this thesis. It was developed to more accurately predict binding energies. The scoring functions have the following form:

$$XP\ GlideScore = E_{coul} + E_{vdW} + E_{bind} + E_{penalty} \quad (2)$$

$$E_{bind} = E_{hyd_enclosure} + E_{hbnn_motif} + E_{hbcc_motif} + E_{PI} + E_{hb_pair} + E_{phobic_pair} \quad (3)$$

$$E_{penalty} = E_{desolv} + E_{ligand_strain} \quad (4)$$

The terms in equation 3 favor binding, while those presented in equation 4 hinder binding. The $E_{hyd_enclosure}$ term represents an

improved model of hydrophobic interactions, if a ligand is placed in an active-site cavity, as opposed to on the surface of the protein, the lipophilic atoms of the ligand are likely to receive better scores, if they are located in a “hydrophobic pocket” of the protein surrounded by lipophilic protein atoms (hydrophobic enclosure), scores should be better than in a location surrounded primarily by polar or charged groups (see Figure 3), the proper treatment of hydrophobic enclosure is the key to discrimination of highly and weakly potent binding motifs and compounds; the $E_{hb_nm_motif}$ term captures hydrogen bonds that represent key molecular recognition motifs and they are important in increasing the potency and specificity of medicinal compounds; the $E_{hb_cc_motif}$ term describes special charged-charged hydrogen bond interactions; the EPI term rewarding pi-stacking and pi-cation interactions; the E_{hb_pair} and the E_{phobic_pair} terms are standard ChemScore-like³³ hydrogen bond and lipophilic pair terms; meanwhile the E_{desolv} term water scoring implements a crude explicit water model using a grid-based methodology adding 2.8 Å spheres, approximating water molecules to high scoring docked poses. Finally, the E_{ligand_strain} term refers to contact penalties, considering the rigid-receptor approximation that is made when the ligand has to adjust to fit into an imperfect and

rigid cavity, often adopting high energy, non-ideal torsion angles, the function penalize poses with close internal contacts. Hence, it contains correction terms for properties, such as charged and strongly polar groups adequately exposed to solvent, not typically well included in other scoring functions.

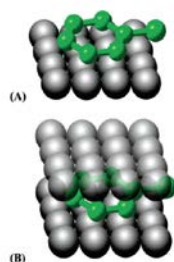


Figure 3 . Ligand interacting with two different environments: (A) a hydrophobic plane and (B) enclosed in a hydrophobic cavity. Taken from Friesner et al Journal of Medicinal Chemistry, 2006, 49, No, 21.

All docking computations within this thesis were performed using the Glide software version 2015-3 from Schrödinger.

2.4. MD simulations

By using Newton's classical equations of motion, computational simulations study the conformational changes of a protein (for instance) with time. MD simulations are being used more and more

to study protein conformational changes due to the importance of generating ensembles of energetically accessible conformations. Molecular Dynamic simulation is a deterministic method based on the temporal evolution of a system. From an initial structure, successive coordinates and velocities are obtained by integrating Newton's equation (5) for the motion in each coordinate direction:

$$\frac{d^2x_i}{dt^2} = \frac{F_i}{m_i} \quad (5)$$

where x_i and m_i are the coordinates and mass of each atom, respectively, t is the time and F_i are the forces acting upon the particles in the system.

The equations are solved simultaneously in small time steps; the coordinates as a function of time represent a trajectory of the system (see Figure 4).

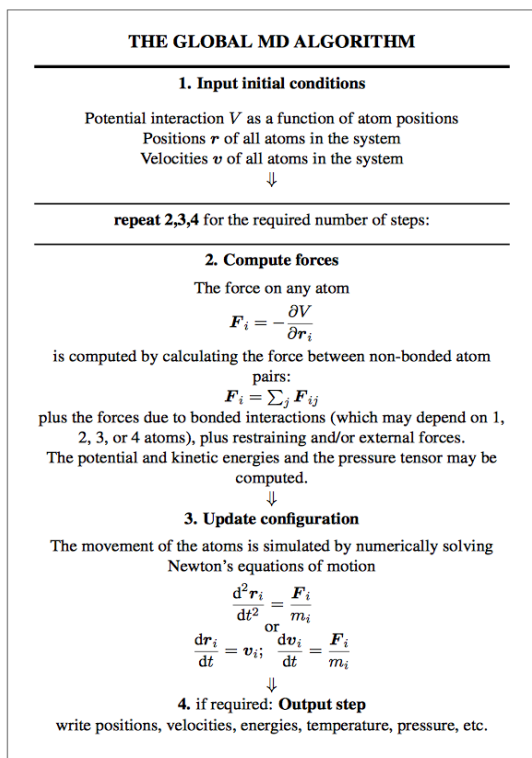


Figure 4 . Global flow scheme for MD simulations. Taken from GROMACS 5.0.7 manual (<ftp://ftp.gromacs.org/pub/manual/manual-5.0.7.pdf>).

There are several algorithms for the integration of the equations of motion, such as Verlet or leap-frog. Due to the large number of particles interacting with each other, the leap-frog algorithm is the most commonly used to update the time step. The leap-frog algorithm is a modification of the Verlet algorithm and its name refers to the fact that the velocity leaps over the coordinate to give the next half-step value of the velocity, which is then used to calculate the new positions. The integration step, the time step, is

limited by the fastest motion in the system. This means typical values of 1fs for atomistic simulations. By restraining bond lengths, using SHAKE³⁴ and LINCS³⁵ algorithms it is possible to use larger time step, typically 2fs.

The simulations performed in this thesis have focused on membrane proteins. The majority of membrane protein X-ray crystallographic structures do not include a membrane environment and at best reveal only a small number of bound lipid or detergent molecules. To fully understand membrane protein function it is essential to accurately insert the protein into a lipid bilayer. Various computational methods have emerged that are able to predict the bilayer-spanning region of a membrane protein structure. These methods permit semi-automatic annotation of membrane protein structures and their results are publicly accessible (e.g. OPM database).³⁶

In order to simulate membrane proteins, different aspects are taken into account. The most relevant are the macroscopic boundary conditions (different ensembles) and different microscopic interaction parameters (force fields).

There are different ensembles such as constant volume (NVT), constant surface tension ($N\gamma T$), and constant isotropic pressure (NPT). Constant volume (NVT), means to keep the dimensions of a box constant, which is the standard condition to simulate a protein in a crystal lattice. However, this condition is not suitable for a membrane protein containing a lipid bilayer, because the dimensions of the box are determined by the area and the length per lipid, which are variable and not well characterized.³⁷ The appropriate boundary condition for a membrane protein that contains a lipid bilayer is therefore constant isotropic pressure (NPT).

The calculation of the interaction energy within a classical description of a molecular system requires a force field. The number of different force fields is even larger than the number of boundary conditions. Beside the packages such as AMBER,¹⁶ CHARMM,³⁸⁻⁴⁰ GROMOS⁴¹, et c. special force fields have been introduced i.e. OPLS, which includes optimized parameters for liquid systems, and more recent versions are optimized also for synthetic small drug-like molecules.

2.4.1. Preparation and execution of a MD simulation

MD simulations were performed with GROMACS v5.0.6.⁴² As a general approach, the receptor complexes were embedded in a pre-equilibrated box (9x9x9 or 10x10x19 nm³) containing a lipid bilayer (205 or 297 POPC molecules) with explicit solvent (~14000 or ~47000 waters) and 0.15 M concentration of Na⁺ and Cl⁻ (~140 or ~490 ions). The exact size, and number of lipids, waters and ions varied depending on the specific system, see each chapter for exact details. Each system was energy minimized and subjected to a 5 step MD equilibration extending 21 ns in total. In the first step (10 ns) the whole system was fixed except hydrogen atoms; in the second step (5 ns), the protein loops were released from restraints; and in the final three steps (2 + 2 + 2 ns) the restraints on the ligand and protein atoms were relaxed from 100, 50 to 10 kJ.mol⁻¹nm⁻², respectively. Unrestrained MD trajectories were produced with varying total time length depending on each system but always using a 2 fs time step. Constant temperature of 300K using separate v-rescale thermostats⁴³ for protein-ligand, lipids, and water plus ions was used. The LINCS algorithm was applied to freeze bond lengths. Lennard-Jones interactions were computed using a 10 Å

cut-off, and the electrostatic interactions were treated using Particle Mesh Ewald (PME)⁴⁴ with a direct sum cut-off of 10 Å. The AMBER99SD-ILDN force field⁴⁵ was used for the protein, the parameters described by Berger et al.³⁷ For lipids, and the general Amber force field (GAFF) and HF/6-31G*-derived RESP atomic charges for the ligand. This combination of protein and lipid parameters has recently been validated.⁴⁶

2.5. Free-energy calculations

A quantitative measure of the favorability of a given (biophysical) process at constant temperature and pressure is the change in Gibbs free-energy (ΔG). Thus, as discussed briefly in the introduction, free-energy calculations are extremely useful for investigating protein-ligand binding affinities or partition coefficients. FE methods provide an estimate of the correct free-energy of some change given a particular set of parameters and physical assumptions.

Relative binding free energies can be calculated using thermodynamics cycles (see Figure 5). All methods for computing free-energy differences consist of the following steps

1. Construct a thermodynamic cycle that allows easy calculation of the free-energy of interest, and determine the end states for each calculation required.
2. Choose a sequence of intermediate states connecting the two end states for each free-energy calculation.
3. Perform equilibrium simulations of the states of interest and any required intermediate states to collect uncorrelated, independent samples.
4. Extract the information of interest required for the desired free-energy method from the sampled configurations.
5. Analyze the information from the simulations to obtain a statistical estimate for the free-energy, including an estimate of statistical error.

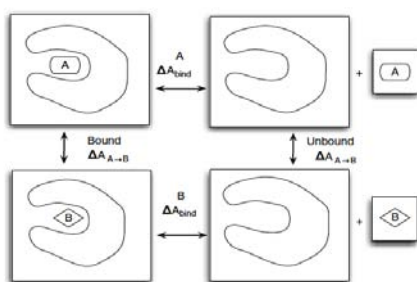


Figure 5. The thermodynamic cycle for the relative binding affinities of ligands A and B to a host molecule. Taken from Biomolecular simulations Methods and Protocols, Springer.31

$$\Delta A_{bind} = \Delta A_{bind}^B - \Delta A_{bind}^A = \Delta A_{A \rightarrow B}^{bound} - \Delta A_{A \rightarrow B}^{unbound} \quad (6)$$

There are two main methodologies to practically construct alchemical pathways between two molecular end states. These are the single topology and the dual topology approach (see Figure 6).^{47,48} In single topology sites correspond to atoms in both molecules. In dual topology no atoms change their type; they merely change from being dummies to fully interacting particles. One advantage to dual topology is that the groups/atoms which change are free to sample the configurational space while being decoupled. This can help increase the sampling, however, more atoms must be annihilated or decoupled from the environment, therefore requiring more intermediates. Despite such large molecular changes, convergence time is often the limiting factor, so a dual topology approach can be more efficient. Single topology is used within the LOMAP⁴⁹ and GR OMACS FEP implementation for instance. Recent reports have highlighted difficulties for ring breaking alchemical modifications arising due to the multiply connected dummy atoms that interact with the remaining system.⁵⁰ This does not occur with dual topology which instead annihilates and replaces a larger ring with a smaller one.

Most of the common simulation packages can perform this kind of calculation (AMBER, CHARMM, GROMACS, Desmond...)

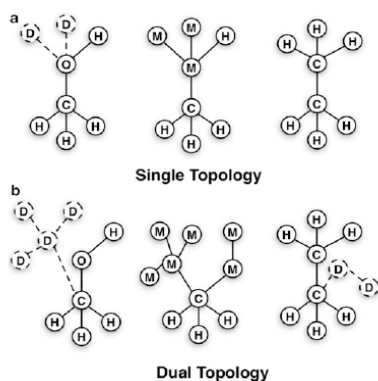


Figure 6 . Schematic example of the two approaches to construct all chemical pathways. Taken from Biomolecular Simulations Methods and Protocols, Springer.

The FEP calculations have been performed using Schrödinger Maestro software. All calculations were conducted using version 2015-3 of the Schrödinger molecular modeling suite. The FEP methodology used here combines an accurate modern force field, OPLSv3 (with parameterization for each ligand calculated up front)⁵¹, the efficient GPU-enabled parallel molecular dynamics engine Desmond version 3.9, the REST enhanced sampling technique^{52,53} and the cycle-closure correction algorithm⁵⁴ to incorporate redundant information into free-energy estimates, it is often referred to as FEP+. Calculations were conducted using the

FEP+ mapper technology to automate setup and analysis. Desmond uses soft core potentials to overcome possible van der Waals endpoint instabilities at $\lambda = 0$ and 1 . Overall default computation protocols were used with a 5 ns simulation length for ligands both in complex and in solution.

We report theoretical error estimates based on cycle closure methodology, the theoretical FEP+ predicted error and also the mean unsigned error compared to experiment. The cycle closure error assesses the reliability of the predictions by determining how much the sum of the calculated free-energy changes, for each closed thermodynamic cycle within the FEP+ mapper, deviates from the theoretical value of 0.⁵⁴ The FEP+ theoretical error (Bennett error) is derived from the Bennett acceptance ratio (BAR) analytical error as the square root of the estimated variance of the total free-energy.^{54,55 56}

2.6. References

- 1 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
- 2 Mason, J. S. *et al.* High end GPCR design: crafted ligand design and druggability analysis using protein structure, lipophilic hot spots and explicit water networks. *In Silico Pharmacology* **1**, 1-12, (2013).
- 3 Rasmussen, S. G. F. *et al.* Crystal structure of the [bgr]2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549-555, (2011).
- 4 Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the [bgr]2 adrenoceptor. *Nature* **469**, 175-180, (2011).
- 5 Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561-567, (2015).
- 6 Tautermann, C. S., Seeliger, D. & Kriegl, J. M. What can we learn from molecular dynamics simulations for GPCR drug design? *Computational and Structural Biotechnology Journal* **13**, 111-121, (2015).
- 7 Cavasotto, C. N. & Pataki, S. S. Homology modeling in drug discovery: current trends and applications. *Drug Discovery Today* **14**, 676-683, (2009).
- 8 Šali, A. & Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **234**, 779-815, (1993).
- 9 Molecular Operating Environment (MOE), 2015.10 (Chemical Computing Group Inc, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7,).
- 10 Knox, A. J. S., Meegan, M. J., Carta, G. & Lloyd, D. G. Considerations in Compound Database Preparation“Hidden” Impact on Virtual Screening Results. *Journal of Chemical Information and Modeling* **45**, 1908-1919, (2005).
- 11 Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27**, 221-234, (2013).

- 12 Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry* **97**, 10269-10280, (1993).
- 13 Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009.).
- 14 Maestro (120 West 45th Street 17th Floor, Tower 45 New York, NY 10036-4041, USA).
- 15 Friesner, R. A. *et al.* Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *Journal of Medicinal Chemistry* **49**, 6177-6196, (2006).
- 16 Case, D. *et al.* *Amber 11*.
- 17 Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* **7**, 2284-2295, (2011).
- 18 Cross, J. B. *et al.* Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *Journal of Chemical Information and Modeling* **49**, 1455-1474, (2009).
- 19 Leach, A. R., Shoichet, B. K. & Peishoff, C. E. Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *Journal of Medicinal Chemistry* **49**, 5851-5855, (2006).
- 20 Chen, H., Lyne, P. D., Giordanetto, F., Lovell, T. & Li, J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *Journal of Chemical Information and Modeling* **46**, 401-415, (2006).
- 21 Sousa, S. F. R., A. J.M.; Coimbra, J.T.S.; Neves, R. P.P.; Martins, S.A.; Moorthy, N.S.H.N.; Fernandes, P.A.; Ramos, M.J. Protein-Ligand Docking in the New Millennium A Retrospective of 10 Years in the Field. *Current Medicinal Chemistry* **20**, 2296-2314 (2013).
- 22 Huang, S.-Y., Grinter, S. Z. & Zou, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics* **12**, 12899-12908, (2010).

- 23 Warren, G . L. *et al.* A Critical Assessment of Docking Programs and Scoring Functions. *Journal of Medicinal Chemistry* **49**, 5912-5931, (2006).
- 24 Sliwoski, G ., Kothiwale, S ., Miesler, J . & Lowe, E . W . Computational Methods in Drug Discovery. *Pharmacological Reviews* **66**, 334 (2013).
- 25 Jorgensen, W . L. The Many Roles of Computation in Drug Discovery. *Science* **303**, 1813 (2004).
- 26 Li, Y ., Han, L., Liu, Z . & Wang, R . Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *Journal of Chemical Information and Modeling* **54**, 1717-1736, (2014).
- 27 Schneider, G . Virtual screening: an endless staircase? *Nat Rev Drug Discov* **9**, 273-276 (2010).
- 28 Repasky, M . P . *et al.* Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *Journal of Computer-Aided Molecular Design* **26**, 787-799, (2012).
- 29 Corbeil, C. R., Williams, C. I. & Labute, P. Variability in docking success rates due to dataset preparation. *Journal of Computer-Aided Molecular Design* **26**, 775-786, (2012).
- 30 Jorgensen, W . L., Maxwell, D . S. & Tirado-Rives, J . Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118**, 11225-11236, (1996).
- 31 Friesner, R. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *Journal of Medicinal Chemistry* **47**, 1739-1749, (2004).
- 32 Halgren, T. A. *et al.* Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **47**, 1750-1759, (2004).
- 33 Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function, and Bioinformatics* **52**, 609-623, (2003).

- 34 Ryckaert, J. -P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **23**, 327-341, (1977).
- 35 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463-1472, (1997).
- 36 Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Research* **40**, D370-D376, (2012).
- 37 Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **72**, 2002-2013 (1997).
- 38 Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187-217, (1983).
- 39 Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **25**, 1584-1604, (2004).
- 40 MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B* **102**, 3586-3616, (1998).
- 41 Scott, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **103**, 3596-3607, (1999).
- 42 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, (2015).
- 43 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126**, 014101, (2007).
- 44 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092, (1993).

- 45 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **78**, 1950-1958, (2010).
- 46 Cordero, A., Caltabiano, G. & Pardo, L. Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters. *Journal of Chemical Theory and Computation* **8**, 948-958, (2012).
- 47 Axelsen, P. H. & Li, D. Improved convergence in dual-topology free energy calculations through use of harmonic restraints. *Journal of Computational Chemistry* **19**, 1278-1283, (1998).
- 48 Michael, R. S. Hirts, D. L. M. *Biomolecular Simulations Methods and Protocols*. 271-311 (Springer New York Heidelberg Dordrecht London, 2013).
- 49 Liu, S. *et al.* Lead optimization mapper: automating free energy calculations for lead optimization. *Journal of Computer-Aided Molecular Design* **27**, 755-770, (2013).
- 50 Liu, S., Wang, L. & Mobley, D. L. Is Ring Breathing Feasible in Relative Binding Free Energy Calculations? *Journal of Chemical Information and Modeling* **55**, 727-735, (2015).
- 51 Shivakumar, D., Harder, E., Damm, W., Friesner, R. A. & Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *Journal of Chemical Theory and Computation* **8**, 2553-2558, (2012).
- 52 Liu, P., Kim, B., Friesner, R. A. & Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13749-13754 (2005).
- 53 Wang, L., Friesner, R. A. & Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **115**, 9431-9438, (2011).
- 54 Wang, L. *et al.* Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *Journal of Chemical Theory and Computation* **9**, 1282-1293, (2013).

- 55 Paliwal, H. & Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *Journal of Chemical Theory and Computation* **7**, 4115-4134, (2011).
- 56 Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **22**, 245-268, (1976).

3. Application of Molecular Dynamics simulations

3.1. Introduction to G protein-coupled receptors

G protein-coupled receptors (GPCRs) are the largest class of membrane proteins in the human genome.¹ GPCRs are able to detect and transduce chemical signals present in the external environment of the cell to the cytoplasmic side (Figure 1). GPCRs transduce sensory signals of external origin such as photons, odors or pheromones and endogenous signals, including neurotransmitters, (neuro)peptides, proteases, glycoprotein hormones and ions.

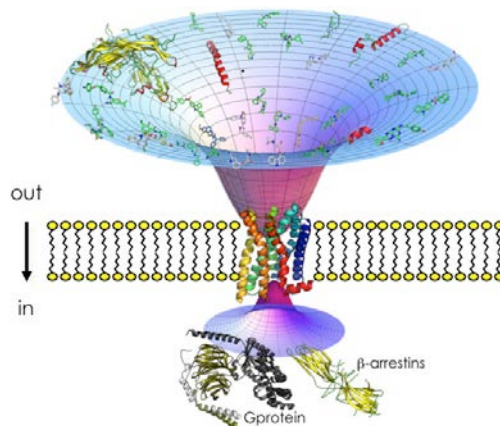


Figure 1 . Illustration of GPCR function, recognizing many different types of extra-cellular signaling molecules such as peptides, neurotransmitters, and hormones and transmitting this signal to the intracellular side of the membrane via activation of the 7-transmembrane helices and the initiation of intracellular signaling cascades through G-proteins and β -arrestin.

The response is operated through secondary messenger cascades controlled by different heterotrimeric guanine nucleotide-binding proteins (G-proteins) coupled and G-protein independent at their intracellular regions (Figure 2).

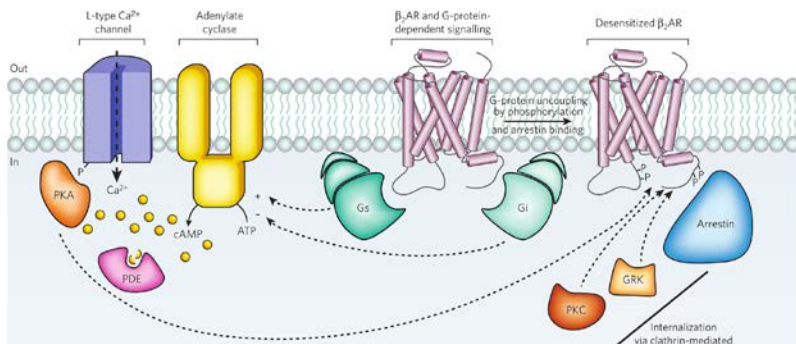


Figure 2 . Schematic showing a typical GPCR interacting with intracellular signaling proteins such as G-proteins or arrestin. Adapted from Dupont et al. Physiology 2010, 25, 293-303.

Due to their relevance to cellular physiology² and their accessibility from the extracellular environment, membrane proteins represent a significant portion of therapeutic drug targets.^{3,4} Reflecting the vast variety of signal impulses, GPCRs have evolved as a very diverse protein superfamily which can be grouped into six classes based on sequence homology and functional similarity: Class A (Rhodopsin-like), Class B (Secretin receptor family), Class C (Metabotropic glutamate/pheromone), Class D (Fungal mating

pheromone receptors), Class E (Cyclic AMP receptors), and Class F (Frizzled/Smoothed).⁵⁻⁷ With more than 700 members, the rhodopsin-like Class-A family is by far the biggest of all GPCR families.

3.2. The structure of G protein-coupled receptors

For many years the only available high resolution crystal structure available for GPCRs was of bovine rhodopsin solved in 2000.⁸ By 2007-2008 advances in experimental crystallization and other techniques permitted solving the crystal structure of other members of class-A GPCRs, including β 2 adrenergic receptor⁹ and the A_{2A} adenosine receptor.¹⁰ To date, 142 GPCR crystal structures (81 of unique ligand-receptor complexes and 37 of unique receptors) are deposited in the Protein Data Bank¹¹. Available crystal structures include receptors from classes A, B,¹² C,¹³ and F,¹⁴ in complex with agonists, antagonists, inverse agonists,¹⁵ allosteric modulators,¹⁶ or biased ligands,¹⁷ in complex with a G protein¹⁸ or with β -arrestin,¹⁹ and in the form of monomers of homo-oligomers (11–15).^{14,20-23} Interestingly, despite their overall low sequence identity,²⁴ GPCRs display a highly conserved molecular

architecture. This architecture is characterized by the presence of seven α -helical transmembrane (7 TM) segments, which span the cell membrane, connected to each other by three extracellular (ECL) and three intracellular (ICL) loops, and a disulfide bridge between ECL2 and TM3 (Figure 3). The N-terminal region is located towards the extracellular side of the membrane whereas the C-terminal region, containing a short α -helix lying perpendicular to the membrane plane (Hx8), faces the intracellular milieu.

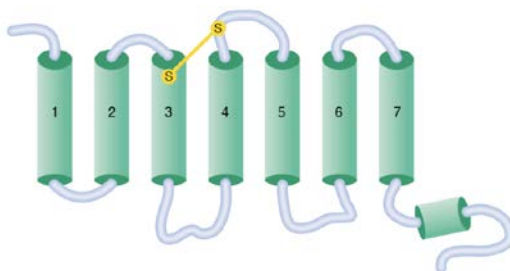


Figure 3. Schematic representation of the 7TM architecture.

Analysis of the known crystal structures of GPCRs shows that ligand binding mostly occurs in a main cavity located between the extracellular segments of TMs 3, 5, 6, and 7. However, reflecting the large variety of stimuli that can interact with GPCRs, the cavity's size and depth within the TM bundle can alter largely between different receptor subfamilies (Figure 4).²⁵

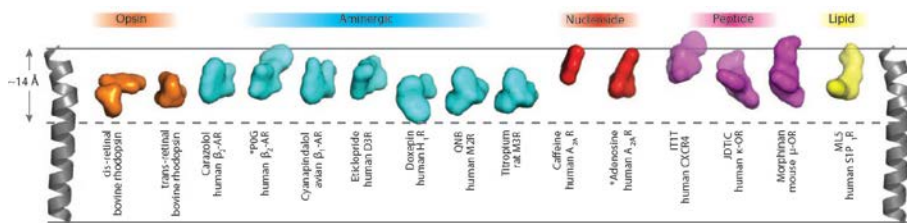


Figure 4: Comparison of size and depth of ligand binding pockets using TM4 as a frame of reference. Figure from Venkatakrisnan et al Nature. 2013, 494, 185–194)

3.3. GPCR activation

GPCR activation involves the binding of ligands in the extracellular part of the TM region, which results in several conformational changes in the TM core. It is now accepted that despite the structurally diverse type of extracellular signals, the processes that propagate the signal from the ligand binding site to the intracellular amino acids of the TM bundle converge towards structurally conserved mechanisms of effector activation.²⁶ While the first steps of these mechanisms will be specific for each subfamily, the final steps will share many common structural features. For instance a conserved ‘transmission switch’²⁷ has been proposed based on the fact that a hydrogen bond interaction between agonists and TM5 in β_1 - and β_2 -receptors or the conformational change of inactive 11-cis retinal to the active 11-

trans retinal in rhodopsin or a agonist-binding to the A_{2A} receptor, stabilizes a receptor conformation that includes an inward movement of TM5 at the highly conserved P^{5.50}, relative to inactive structures. This inward movement of TM5 is sterically competing with a bulky hydrophobic side chain at position 3.40, triggering a small anticlockwise rotation, viewed from the extracellular side, of TM3. This rotation of TM3 repositions the side chain of F^{6.44}, facilitating the outward movement of TM6 for receptor activation and G protein binding. This ‘transmission switch’ is also present in family B GPCRs^{28,29}, and other intracellular conformational switches have been identified which are important for receptor activation (Figure 5).

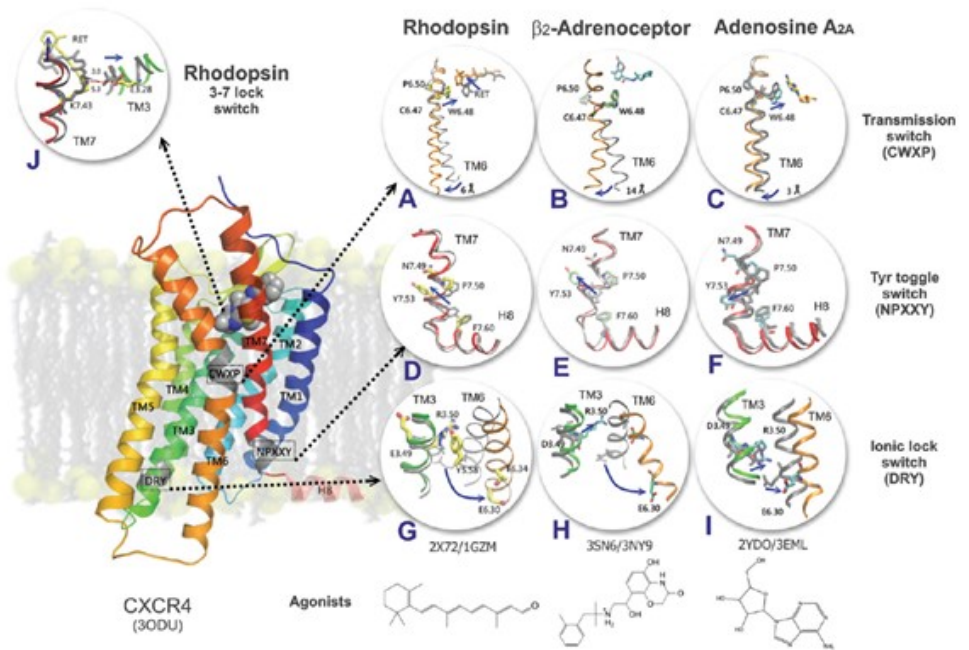


Figure 5. Action of molecular switches in GPCRs. Taken from Trzaskowski et al. *Current Medicinal Chemistry*, 2012, 19, 1090-1109.

Either direct ligand interaction or allosteric interference with these switches is likely a source of small molecule induced modification of the conformational behavior of the receptor, shifting the balance from active to inactive conformational states.³⁰ These small local structural changes near the binding site are translated into larger-scale helix movements at the intracellular site, mainly TMs 5 and 6, opening a cavity for the binding of the C-terminal $\alpha 5$ helix of the G-protein.¹⁸

3.4. Allosteric modulators

Traditionally most small molecules drugs have acted at the orthosteric GPCR binding site where they bind competitively with the endogenous agonist, (Figure 6). However, increasing attention is turning towards the advantageous approach of allosteric modulation of GPCRs. The Greek root for allo- means another. Allosteric literally translated from its Greek root means 'other site'. In contrast with orthosteric ligands, allosteric modulators of GPCRs interact with binding sites different than endogenous ligands. Allosteric modulators do not activate receptors on their own, but modulate the response in the presence of the endogenous ligand. Based on their effects, allosteric modulators are organized into three different groups:

Positive Allosteric Modulators (PAMs): enhance the affinity and/or responsiveness of the orthosteric ligand.

Negative Allosteric Modulators (NAMs): decrease the affinity and/or responsiveness of the orthosteric ligand.

Neutral Allosteric Modulators (also referred to as Silent Allosteric Modulators: SAMs): do not alter the affinity or efficacy of the

orthosteric ligand, but they provide important tools to validate binding sites.

The discovery of allosteric ligands has enriched the ways in which the functions of GPCRs can be manipulated for potential therapeutic benefit. In general, the orthosteric binding site is highly conserved within a GPCR family due to the evolutionary pressure to retain amino acid sequences necessary for binding of the endogenous ligand. In contrast, allosteric sites are often less conserved, and hence can assist greatly with the identification of selective small molecule modulators, which is often desirable for the purpose of biological validation or therapeutic purposes. Also, orthosteric binding sites may have inherent undesirable properties which can be avoided by targeting a distinct site, for instance the high polarity of the glutamate binding site in metabotropic glutamate (mGlu) receptors is detrimental for CNS-targeting therapeutics, or the peptide binding sites of class B GPCRs may be detrimental for identifying drug-like small molecules. Also, endogenous agonists are often synthesized by the body, *in situ*, and on demand, producing transient receptor stimulation and cellular activation. Traditional orthosteric drugs usually have a very different profile, possibly in an undesirable way causing target-

related toxic effects and receptor desensitization. Allosteric modulators only function in the presence of the endogenous agonist and as such should respond to and work in harmony with the physiological changes in a agonist. Therefore allosteric modulators are less likely to cause receptor desensitization.³¹⁻³³ There are currently only two marketed GPCR allosteric modulators: Cinacalcet and Maraviroc, which target the calcium-sensing receptor and chemokine CCR5, respectively.

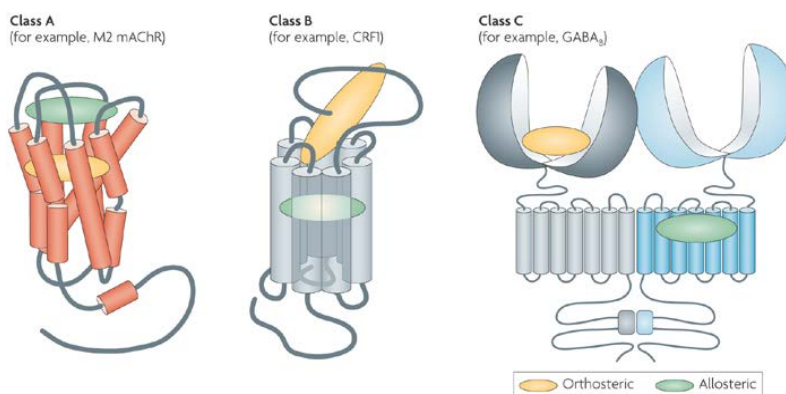


Figure 6. Schematic of GPCR structures showing orthosteric and allosteric binding sites. Taken from Conn, P. J.; Christopoulos, A.; Lindsay, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat Rev Drug Discov* 2009, 8, 41-54.

3.5. Oligomerization of GPCRs

GPCRs have been classically described as monomeric transmembrane receptors that form a ternary complex: a ligand, the GPCR, and its associated G-protein. Thus, conventional drug design targeting GPCRs has mainly focused on the inhibition/activation of a single receptor. Nevertheless, it is now well accepted that many GPCRs oligomerize in cells.³⁴ There is a broad consensus that Class C receptors, in particular, exist as stable dimers, both as homodimeric receptors, for example the covalently bound metabotropic glutamate receptor (mGlu₅) homodimer, and as heterodimeric receptors, as seen in the GABA_B receptor 1 and GABA_B receptor 2, heterodimer. A number of other GPCRs have also shown to dimerize, most notably the adrenergic receptors, opioid receptors, somatostatin receptors and other druggable targets.³⁵ New combinations of receptors are continually being discovered to form homo- and heterodimers, as well as higher order oligomers in natural tissues like A_{2A}R/D₂R/CB₁³⁶ or A₁R/A_{2A}R.³⁷

In addition, receptor activation is modulated by allosteric communication between protomers of class A GPCR dimers.³⁸ The minimal signaling unit, two receptors and a single G protein, is

maximally activated by agonist binding to a single protomer. Inverse agonist binding to the second protomer enhances signaling, whereas an agonist binding to the second protomer blunts signaling. Thus, GPCR dimer function can be modulated by the activity state of the second protomer, which for a heterodimer may be altered in pathological states.

Cysteine cross-linking experiments have suggested that receptor oligomerization involves the surfaces of TM1, 4, and/or 5.³⁹ Recently, the crystal structure of the μ -opioid receptor has revealed crystallographic two-fold axis through the TM1 and TM5 interfaces,⁴⁰ while the crystal structure of the histamine H₁-receptor shows a parallel dimer through TM4.⁴¹ (Figure 7)

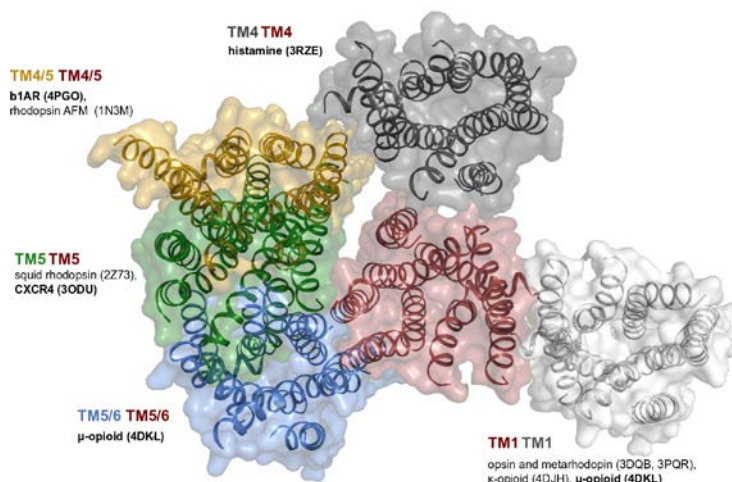


Figure 7. Crystallographic interface of several GPCRs. Adapted from Cordero et al. Trends in Biochemical Sciences, 2015, 40, 548-551.

The metabotropic glutamate (mGlu) receptors are family C GPCRs that participate in the modulation of synaptic transmission and neuronal excitability throughout the central nervous system (CNS). These receptors are distinguished from their family A relatives by the presence of a large extracellular N-terminal domain, (Figure 8). Known as the Venus fly trap domain (VFD) this domain contains the endogenous ligand-binding site, the glutamate-binding site.⁴² Evidence suggests that two VFDs dimerize together, back to back, and large conformational changes are induced when a agonist binds to one or both VFDs.⁴³ These conformational changes induced by ligand binding are propagated from the VFD via cysteine-rich domains (CRDs) which connects the 7TM. The CRD contains nine critical cysteine residues, eight of which are linked by disulfide bridges.⁴⁴ Significant effort has been made to understand the large scale domain movements involved in receptor activation. Activation of the mGlu₂ receptor shifts the 7-TM dimer interface from TMs 4 and 5 in the inactive state to TM6 in the active state.⁴⁵

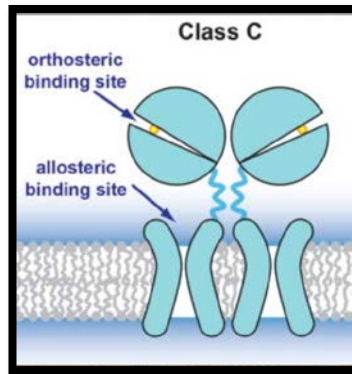


Figure 8. Schematic representation of the structure of Class C GPCRs. Adapted from Wu et al. *Science*, 2014, 344, 58-64.

Family C GPCRs also include GABA_B receptors, calcium-sensing receptors, pheromone receptors and taste receptors.⁴² The mGlu receptors are subclassified into three groups based on sequence homology, G-protein coupling, and ligand selectivity (Table 1).

Group I	Group II	Group II
mGlu ₁ , mGlu ₅	mGlu ₂ , mGlu ₃	mGlu ₄ , mGlu ₆ , mGlu ₇ , mGlu ₈
Gq	Gi/o	Gi/o
↑IP3/DAG, ↑se Ca ²⁺	↓se cAMP	↓se cAMP
Somatodendric	Somatodendric and nerve terminals	Nerve terminals

Table 1. Classification of mGlu receptors according to sequence similarity (first row), coupling G-protein, intracellular effect, and location.

Because families A and C of GPCRs bind orthosteric ligands and allosteric modulators, respectively, within binding pockets located in a similar position in the 7TM domain (Figure 9)⁴⁶⁻⁵⁰ and maintain the spatial conservation of the TM helices, despite the low sequence identity, we may expect similar mechanisms of receptor activation.

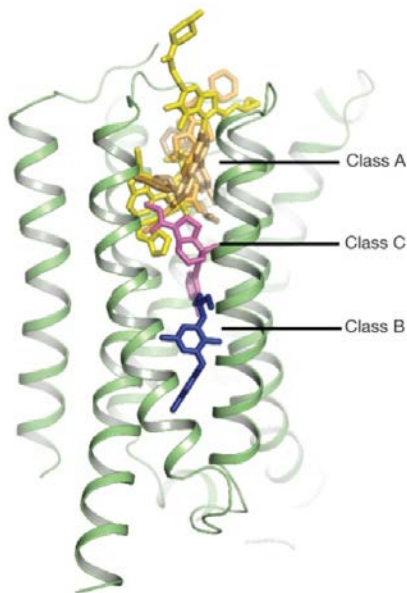


Figure 9. Comparison of the orthosteric binding site of Class A and B with the allosteric binding site of Class C. Shown are the Class C NAM mavoglurant (magenta), an antagonist bound to CFF₁R (blue) and a selection of Class A ligands (yellow). Adapted from Doré et al. Nature 2014, 511, 557-562.

The recent Class C crystal structures of mGlu₁ and mGlu₅ receptors solved with NAMs have confirmed the 7-TM binding site.^{13,51} Crystal structures are also available of the extracellular

domains of several mGlu receptors such as mGlu₁⁵² and mGlu₂.⁵³ However, there is no full length crystal structure of the entire protein, in either monomeric or dimeric state.

3.6. mGlu2 receptor as a drug target

Glutamate is the major excitatory neurotransmitter in the central nervous system (CNS) of vertebrates and elicits and modulates synaptic responses in the CNS via ionotropic and metabotropic glutamate receptors (mGlu receptors). Glutamate plays a pivotal role in numerous physiological functions, such as learning and memory but also sensory perception, development of synaptic plasticity, motor control, respiration, and regulation of cardiovascular function.⁵⁴ An imbalance in glutamatergic neurotransmission is believed to be at the center of various neurological and psychiatric diseases.⁵⁵⁻⁶¹ The mGlu receptors contribute to the fine-tuning of synaptic efficacy and can be considered slow responders to glutamate effect, compared to faster ionotropic channels.^{54,62}

Of the eight mGlu receptors the mGlu₂ receptor has proven to be of importance in neuropharmacology. Preferentially expressed on

presynaptic nerve terminals, mGlu2 receptors negatively modulate glutamate and GABA release and are widely distributed in the brain.⁶³ High levels of mGlu2 receptor are seen in brain areas such as prefrontal cortex, hippocampus and amygdala where glutamate hyperfunction may be implicated in disorders and diseases such as anxiety and schizophrenia.⁶⁴⁻⁶⁷ It is therefore considered that activation of group II mGlu (Table 1) receptors may offer anxiolytic and/or antipsychotic effects.⁶⁸

The first generation of mGlu2 activators were conformationally constrained agonists, analogs of glutamate.⁶⁹ Examples showed efficacy in preclinical models driven by excessive glutamate transmission, including stress/anxiety, pain and psychosis. Meanwhile clinical effects were seen in both generalized anxiety disorder and schizophrenia patients. Although the findings were not replicated in larger studies, analyses suggest subgroups were responsive to the mechanism of action.

Given the promise offered by mGlu2/3 receptor activation there has been considerable efforts to identify novel drug-like mGlu2 receptor positive allosteric modulators.⁷⁰⁻⁷² Like agonists, such molecules have also shown efficacy in preclinical models of LY354740, LY544344 and LY379268. Two examples are known to

have advanced to clinical trials, JNJ-40411813 and AZD8529.⁷³ The former did not show efficacy in patients with major depressive disorder with significant anxiety symptoms, but in a Phase II study in schizophrenia, it met the primary objectives of safety and tolerability and also demonstrated an effect in patients with residual negative symptoms.⁷⁴ Meanwhile the latter compound from AstraZeneca was discontinued in a phase II trial in schizophrenic patients for unknown reasons.

Alternatively, blockade of mGlu2/3 receptors is a novel approach to enhance glutamate transmission, having potential therapeutic benefit in CNS disorders for which diminished glutamate transmission is suggested. In this regard, mGlu2/3 antagonists elicit robust antidepressant-like behaviors,⁷⁵⁻⁷⁹ cognitive enhancing effects⁸⁰⁻⁸² and promote wakefulness⁸³ in rodents. At least two mGlu2/3 antagonists have entered clinic, an orthosteric antagonist oral prodrug BCI-838 which is a prodrug for BCI-632, and a negative allosteric modulator (NAM), RO4995819, RG1538, Decoglurant.⁸⁴ The latter molecule showed a lack of efficacy on inventories of depression, cognition and physical functioning when dosed in conjunction with either selective serotonin uptake or mixed

serotonin/norepinephrine uptake inhibitors in non-responding depressed patients.⁸⁵

In conclusion there is still significant pharmaceutical interest in discovering and further exploring positive and negative allosteric modulators of mGlu2 receptors for indications in CNS disorders.

3.7. References

- 1 Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *PROTEOMICS* **10**, 1141-1149, (2010).
- 2 Smit, M. J. *et al.* Pharmacogenomic and Structural Analysis of Constitutive G Protein-Coupled Receptor Activity. *Annual Review of Pharmacology and Toxicology* **47**, 53-87, (2007).
- 3 Arinaminpathy, Y., Kurana, E., Engelman, D. M. & Gerstein, M. B. Computational analysis of membrane proteins: the largest class of drug targets. *Drug discovery today* **14**, 1130-1135, (2009).
- 4 Imming, P., Sinning, C. & Meyer, A. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* **5**, 821-834 (2006).
- 5 Bjarnadóttir, T. K. *et al.* Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics* **88**, 263-273, (2006).
- 6 Lagerstrom, M. C. & Schiöth, H. B. Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* **7**, 339-357 (2008).
- 7 Fredriksson, R., Lagerström, M. C., Lundin, L.-G. & Schiöth, H. B. The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Molecular Pharmacology* **63**, 1256 (2003).
- 8 Palczewski, K. *et al.* Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* **289** (2000).
- 9 Cherezov, V. *et al.* High Resolution Crystal Structure of an Engineered Human β_2 -Adrenergic G Protein-Coupled Receptor *Science* **318**, 1258 (2007).
- 10 Jaakola, V.-P. *et al.* The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist. *Science* **322**, 1211 (2008).
- 11 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
- 12 Hollenstein, K. *et al.* Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* **499**, 438-443, (2013).

- 13 Wu, H. *et al.* Structure of a Class C GPCR Metabotropic Glutamate Receptor 1 Bound to an Allosteric Modulator. *Science* **344**, 58 (2014).
- 14 Wang, C. *et al.* Structure of the human smoothed receptor 7TM bound to an antitumor agent. *Nature* **497**, 338 -343, (2013).
- 15 Kobilka, B. K. & Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends in Pharmaceutical Sciences* **28**, 397-406, (2007).
- 16 Christopher, J. A. *et al.* Fragment and Structure-Based Drug Discovery for a Class C GPCR: Discovery of the mGlu5 Negative Allosteric Modulator HTL14242 (3-Chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzotrile). *Journal of Medicinal Chemistry* **58**, 6653-6664 (2015).
- 17 Saulière, A. *et al.* Deciphering biased-agonism complexity reveals a new active AT1 receptor entity. *Nat Chem Biol* **8**, 622-630, (2012).
- 18 Rasmussen, S. G. F. *et al.* Crystal structure of the [βγ]2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549 -555, (2011).
- 19 Kang, Y. *et al.* Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* **523**, 561 -567, (2015).
- 20 Manglik, A. *et al.* Crystal structure of the μ-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321 -326, (2012).
- 21 Wu, H. *et al.* Structure of the human kappa opioid receptor in complex with JDTic. *Nature* **485**, 327 -332, doi:10.1038/nature10939 (2012).
- 22 Wu, B. *et al.* Structures of the CXCR4 chemokine receptor in complex with small molecule and cyclic peptide antagonists. *Science (New York, N.Y.)* **330**, 1066 -1071, (2010).
- 23 Huang, J., Chen, S., Zhang, J. J. & Huang, X.-Y. Crystal Structure of Oligomeric β(1)-Adrenergic G Protein-Coupled Receptors in Ligand-Free Basal State. *Nature structural & molecular biology* **20**, 419-425, (2013).
- 24 Gonzalez, A., Cordomí, A., Caltabiano, G. & Paro, L. Impact of Helix Irregularities on Sequence Alignment and Homology Modeling of G Protein-Coupled Receptors. *ChemBioChem* **13**, 1393-1399, (2012).

- 25 Venkatakrisnan A J, D. X., Lebon G, Tate C G, Schertler
GF, Babu MM. Molecular signatures of G-protein-coupled
receptors. *Nature* **494**, 185-194 (2013).
- 26 Venkatakrisnan, A. J. *et al.* Diverse activation pathways in
class A G PCRs converge near the G -protein-coupling
region. *Nature* **536**, 484-487, (2016).
- 27 Venkatakrisnan, A. J. *et al.* Molecular signatures of G -
protein-coupled receptors. *Nature* **494**, 185-194, (2013).
- 28 Spyridaki, K. *et al.* Structural-functional analysis of the third
transmembrane domain of the corticotropin-releasing factor
type 1 receptor: role in activation and allosteric antagonism.
J Biol Chem **289**, 18966-18977, (2014).
- 29 Cordomi, A. *et al.* Functional elements of the gastrin
inhibitory polypeptide receptor: Comparison between
secretin- and rhodopsin-like G protein-coupled receptors.
Biochemical pharmacology **96**, 237-246, (2015).
- 30 Manglik, A. *et al.* Structural Insights into the Dynamic
Process of beta2-Adrenergic Receptor Signaling. *Cell* **161**,
1101-1111, (2015).
- 31 Pin, J.-P., Parmentier, M.-L. & Prézeau, L. Positive
Allosteric Modulators for γ -Aminobutyric Acid_B Receptors
Open New Routes for the Development of Drugs Targeting
Family 3 G -Protein-Coupled Receptors. *Molecular
Pharmacology* **60**, 881 (2001).
- 32 Gjoni, T. & Urwyler, S. Receptor activation involving
positive allosteric modulation, unlike full agonism, does not
result in G ABAB receptor desensitization.
Neuropharmacology **55**, 1293-1299, (2008).
- 33 Urwyler, S. Allosteric Modulation of Family C G-Protein-
Coupled Receptors: from Molecular Insights to Therapeutic
Perspectives. *Pharmacological Reviews* **63**, 59 (2011).
- 34 Pin, J. P. *et al.* International Union of Basic and Clinical
Pharmacology. LXVII. Recommendations for the
recognition and nomenclature of G protein-coupled receptor
heteromultimers. *Pharmacol. Rev.* **59**, 5-13 (2007).
- 35 Borroto-Escuela, D. O. *et al.* The G Protein-Coupled
Receptor Heterodimer Network (GPCR-HetNet) and Its Hub
Components. *International Journal of Molecular Sciences*
15, 8570-8590, (2014).
- 36 Navarro, G. *et al.* Interactions between Intracellular
Domains as Key Determinants of the Quaternary Structure

- and Function of Receptor Heteromers. *The Journal of Biological Chemistry* **285**, 27346-27359, (2010).
- 37 Navarro, G. *et al.* Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with G(i) and G(s). *BMC Biology* **14**, 26, (2016).
- 38 Han, Y., Moreira, I. S., Urizar, E., Weinstein, H. & Javitch, J. A. Allosteric communication between protomers of dopamine class A GPCR dimers modulates activation. *Nat. Chem. Biol.* **5**, 688-695, (2009).
- 39 Guo, W. *et al.* Dopamine D2 receptors form higher order oligomers at physiological expression levels. *EMBO J.* **27**, 2293-2304 (2008).
- 40 Manglik, A. *et al.* Crystal structure of the micro-opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321-326, (2012).
- 41 Shimamura, T. *et al.* Structure of the human histamine H1 receptor complex with doxepin. *Nature* **475**, 65-70, doi:10.1038/nature10236 (2011).
- 42 Pin, J.-P., Galvez, T. & Prézeau, L. Evolution, structure, and activation mechanism of family 3 /C G-protein-coupled receptors. *Pharmacology & Therapeutics* **98**, 325-354, (2003).
- 43 Jingami, H., Nakanishi, S. & Morikawa, K. Structure of the metabotropic glutamate receptor. *Current Opinion in Neurobiology* **13**, 271-278, (2003).
- 44 Muto, T., Tsuchiya, D., Morikawa, K. & Jingami, H. Structures of the extracellular regions of the group II/III metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences* **104**, 3759-3764 (2007).
- 45 Xue, L. *et al.* Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nat Chem Biol* **11**, 134-140, (2015).
- 46 Farinha, A. *et al.* Molecular determinants of positive allosteric modulation of the human metabotropic glutamate receptor 2. *British Journal of Pharmacology* **172**, 2383-2396, doi:10.1111/bph.13065 (2015).
- 47 Hemstapat, K. *et al.* A Novel Family of Potent Negative Allosteric Modulators of Group II Metabotropic Glutamate Receptors. *Journal of Pharmacology and Experimental Therapeutics* **322**, 254 (2007).

- 48 Lundström, L . *et al.* Structural determinants of allosteric antagonism at the metabotropic glutamate receptor 2 : mechanistic studies with new potent negative allosteric modulators. *British Journal of Pharmacology* **164**, 521-537, (2011).
- 49 Rowe, B . A . *et al.* Transposition of Three Amino Acids Transforms the Human Metabotropic Glutamate Receptor (mGluR)-3-Positive Allosteric Modulation Site to mGluR2, and Additional Characterization of the mGluR2-Positive Allosteric Modulation Site. *Journal of Pharmacology and Experimental Therapeutics* **326**, 240 (2008).
- 50 Schaffhauser, H . *et al.* Pharmacological Characterization and Identification of Amino Acids Involved in the Positive Modulation of Metabotropic Glutamate Receptor Subtype 2. *Molecular Pharmacology* **64**, 798 (2003).
- 51 Dore, A. S. *et al.* Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**, 557-562, (2014).
- 52 Kunishima, N . *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971-977 (2000).
- 53 Monn, J . A . *et al.* Synthesis and Pharmacological Characterization of C-4-(Thiotriazolyl)-substituted-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylates. Identification of (1R,2S,4R,5R,6R)-2-Amino-4-(1H-1,2,4-triazol-3-ylsulfanyl)bicyclo[3.1.0]hexane-2,6-dicarboxylic Acid (LY2812223), a Highly Potent, Functionally Selective mGlu2 Receptor Agonist. *Journal of Medicinal Chemistry* **58**, 7526-7548, (2015).
- 54 Watkins, J. C. L-Glutamate as a Central Neurotransmitter: Looking Back. *Biochemical Society Transactions* **28**, 297 (2000).
- 55 Niswender, C. M. & Conn, P. J. Metabotropic Glutamate Receptors: Physiology, Pharmacology, and Disease. *Annual review of pharmacology and toxicology* **50**, 295-322, (2010).
- 56 Pile, A ., Chaki, S ., Nowak, G . & Witkin, J. M . Mood disorders: Regulation by metabotropic glutamate receptors. *Biochemical Pharmacology* **75**, 997-1006, (2008).
- 57 Lüscher, C . & Huber, K. M . Group 1 mGluR-dependent synaptic long-term depression (mGluR-LTD): mechanisms

- and implications for circuitry & disease. *Neuron* **65**, 445-459, (2010).
- 58 Vinson, P. N. & Conn, P. J. Metabotropic Glutamate Receptors as Therapeutic Targets for Schizophrenia. *Neuropharmacology* **62**, 1461-1472, (2012).
- 59 Chiechio, S. & Nicoletti, F. Metabotropic glutamate receptors and the control of chronic pain. *Current Opinion in Pharmacology* **12**, 28-34, (2012).
- 60 Olive, M. F. Metabotropic glutamate receptor ligands as potential therapeutics for addiction. *Current drug abuse reviews* **2**, 83-989 (2009).
- 61 Johnson, K. A., Conn, P. J. & Niswender, C. M. Glutamate receptors as therapeutic targets for Parkinson's disease. *CNS & neurological disorders drug targets* **8**, 475-491 (2009).
- 62 Kew, J. N. C. & Kemp, J. A. Ionotropic and metabotropic glutamate receptors: structure and pharmacology. *Psychopharmacology* **179**, 4-29, (2005).
- 63 Cartmell, J. & Schoepp, D. D. Regulation of Neurotransmitter Release by Metabotropic Glutamate Receptors. *Journal of Neurochemistry* **75**, 889-907, (2000).
- 64 Ohishi, H., Higemoto, R., Akanishi, S. & Mizuno, N. Distribution of the messenger RNA for a metabotropic glutamate receptor, mGluR2, in the central nervous system of the rat. *Neuroscience* **53**, 1009-1018, (1993).
- 65 Richards, G. *et al.* Distribution and abundance of metabotropic glutamate receptor subtype 2 in rat brain revealed by [3H]LY354740 binding in vitro and quantitative radioautography: Correlation with the sites of synthesis, expression, and agonist stimulation of [35S]GTPγS binding. *The Journal of Comparative Neurology* **487**, 15-27, (2005).
- 66 Gu, G. *et al.* Distribution of metabotropic glutamate 2 and 3 receptors in the rat forebrain: Implication in emotional responses and central disinhibition. *Brain Research* **1197**, 47-62, (2008).
- 67 Ghose, S., Gleason, K. A., Potts, B. W., Lewis-Amezcu, K. & Tamminga, C. A. Differential expression of metabotropic glutamate receptor 2 and 3 in schizophrenia: a mechanism for antipsychotic drug action? *The American journal of psychiatry* **166**, 812-820, (2009).

- 68 Shigeyuki, C. & Hirohiko, H. Targeting of Metabotropic Glutamate Receptors for the Treatment of Schizophrenia. *Current Pharmaceutical Design* **17**, 94-102, (2011).
- 69 Monn, J. A. *et al.* Design, Synthesis, and Pharmacological Characterization of (+)-2-Aminobicyclo[3.1.0]hexane-2,6-dicarboxylic Acid (LY354740): A Potent, Selective, and Orally Active Group 2 Metabotropic Glutamate Receptor Agonist Possessing Anticonvulsant and Anxiolytic Properties. *Journal of Medicinal Chemistry* **40**, 528-537, (1997).
- 70 Trabanco, A. A., Cid, J. M., Lavreysen, H., Macdonald, G. J. & Tresadern, G. Progress in the Development of Positive Allosteric Modulators of the Metabotropic Glutamate Receptor 2. *Current Medicinal Chemistry* **18**, 47-68, (2011).
- 71 Trabanco, A. A. & Cid, J. M. mGluR2 positive allosteric modulators: a patent review (2009 – present). *Expert Opinion on Therapeutic Patents* **23**, 629-647, (2013).
- 72 Gyorgy, S. & Gyorgy Miklos, K. Positive Allosteric Modulators for mGluR2 Receptors: A Medicinal Chemistry Perspective. *Current Topics in Medicinal Chemistry* **14**, 1771-1788, (2014).
- 73 Litman, R. E. *et al.* AZD8529, a positive allosteric modulator of the mGluR2 receptor, does not improve symptoms in schizophrenia: A proof of principle study. *Schizophrenia Research* **172**, 152-157, (2016).
- 74 Patil ST, Z. L., Martenyi F, Lowe SL, Jackson KA, Andreev BV, Avedisova AS, Bardenstein LM, Gurovich IY, Morozova MA, Mosolov SN, Neznanov NG, Reznik AM, Smulevich AB, Tochilov VA, Johnson BG, Monn JA, Schoepp DD. Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized Phase 2 clinical trial. *Nature medicine* **13**, 1102-1107 (2007).
- 75 Chaki, S. *et al.* MGS0039: a potent and selective group II metabotropic glutamate receptor antagonist with antidepressant-like activity. *Neuropharmacology* **46**, 457-467, (2004).
- 76 Chaki, S., Ago, Y., Palucha-Paniewiera, A., Matrisciano, F. & Pilc, A. mGlu2/3 and mGlu5 receptors: Potential targets for novel antidepressants. *Neuropharmacology* **66**, 40-52, (2013).

- 77 Yoshimizu, T ., Shimazaki, T ., Ito, A . & Chaki, S . A n mGluR2/3 antagonist, MGS0039, exerts antidepressant and anxiolytic effects in behavioral models in rats. *Psychopharmacology* **186**, 587, doi :10.1007/s00213-006-0390-7 (2006).
- 78 Ago, Y . *et al.* The selective metabotropic glutamate 2/3 receptor antagonist MGS0028 reverses psychomotor abnormalities and recognition memory deficits in mice lacking the pituitary adenylate cyclase-activating polypeptide. *Behavioural Pharmacology* **24** (2013).
- 79 Pałucha-Poniewiera, A . *et al.* On the mechanism of the antidepressant-like action of group II mGlu receptor antagonist, MGS0039. *Psychopharmacology* **212**, 523 -535, (2010).
- 80 Woltering, T. J. *et al.* Synthesis and characterization of 1,3-dihydro-benzo[b][1,4]diazepin-2-one derivatives: Part 4. In vivo active potent and selective non-competitive metabotropic glutamate receptor 2/3 antagonists. *Bioorganic & Medicinal Chemistry Letters* **20**, 6969-6974, (2010).
- 81 Goeldner, C . *et al.* Cognitive impairment in major depression and the mGlu2 receptor as a therapeutic target. *Neuropharmacology* **64**, 337-346, (2013).
- 82 Kim, S. H. *et al.* Proneurogenic Group II mGluR antagonist improves learning and reduces anxiety in Alzheimer A β oligomer mouse. *Molecular Psychiatry* **19**, 1235 -1242, (2014).
- 83 Feinberg, I., Schoepp, D .D., Hsieh, K .C., Darchia, N ., Campbell, I .G. The metabotropic glutamate (mGlu)2/3 receptor antagonist LY341495 [2S-2-amino-2-(1S,2S-2-carboxycyclopropyl-1-yl)-3-(xanth-9-yl)propanoic acid] stimulates waking and fast electroencephalogram power and blocks the effects of the mGlu2/3 receptor agonist LY379268 [(-)-2-oxa-4-aminobicyclo[3.1.0]hexane-4,6-dicarboxylate] in rats. *J. Pharmacol. Exp. Ther* **312**, 826-833 (2005).
- 84 Gatti, S ., Goetschi, E ., Palmer, W .S., Wichmann, J ., Woltering, T .J., Preparation of a cetylenyl-pyrazolopyrimidine derivatives for use as mGluR2 antagonists treating CNS disorders. (2006).
- 85 Umbricht, D . *et al.* P.2.f.021 Results of a double-blind placebo-controlled study of the antidepressant effects of the

mGLU2 negative allosteric modulator RG1578. *European Neuropsychopharmacology* **25**, S447,

3.8. Quaternary structure of a G-protein-coupled receptor heterotetramer coupled to G_i and G_s

This Chapter is a collaboration with Rafael Franco, Gemma Navarro, Marc Brugarolas, Estefania Moreno, David Aguinaga, Antoni Cortés, Vicent Casadó, Josefa Mallol, Enric I. Canela, Carme Lluís at the University of Barcelona, Monika Zelman-Femiak at the Max Planck Institute for Intelligent Systems, Peter McCormick at the University of East Anglia and Arnau Corderó and Leonardo Pardo from the Universitat Autònoma de Barcelona. This work is currently published at BMC Biology 2016, 5, 14-26.

3.8.1. Background

G-protein-coupled receptor (GPCR) oligomerization is heavily supported on recent biochemical and structural data.¹⁻⁶ Optical-based techniques are instrumental to study the dynamics and organization of receptor complexes in living cells.⁷ For instance, total internal reflection fluorescence microscopy shows that 30% of muscarinic M₁ receptors exist as dimers (with no evidence of higher oligomers) that undergo interconversion with monomers on a timescale of seconds.⁸ Similarly, the β_1 -adrenergic receptors (β_1 -AR)

are expressed as a mixture of monomers and dimers whereas β_2 -adrenergic receptors (β_2 -AR) have a tendency to form dimers and higher-order oligomers.⁹ Moreover, the monomer-dimer equilibrium of the chemo attractant *N-formyl* peptide receptor at a physiological level of expression lies within a time scale of milliseconds.¹⁰ Together, these studies in heterologous systems show that a given GPCR is present in a dynamic equilibrium between monomers, dimers and higher order oligomers.

Studies in a broad spectrum of GPCRs¹¹⁻¹⁴ show that these receptors may form heteromers. GPCR heteromers are defined as novel signaling units with functional properties different from homomers and represent a completely new field of study.¹⁵ Innovative crystallographic techniques have permitted to obtain crystal structures of families A, B, C and F of GPCRs, bound to either agonists, antagonists, inverse agonists, or allosteric modulators, in the form of monomers or homo-oligomers, in complex with a G protein or with a β -arrestin.¹⁶ Nevertheless, crystal structures of GPCR heteromers have not yet been obtained. Here we propose a quaternary structure of a heteromer, taking into account the molecular stoichiometry and the interacting G proteins.

Adenosine A₁-A_{2A} receptor (A₁R-A_{2A}R) complexes constitute a paradigm in the GPCR heteromer field as A₁R is coupled to G_i and A_{2A}R to G_s, i.e. they transduce opposite signals in cAMP-dependent intracellular cascades. First described as a concentration-sensing device in striatal glutamatergic neurons,¹⁷ the A₁R-A_{2A}R heteromer is thought to function as a G_s/G_i-mediated switching mechanism by which low and high concentrations of adenosine inhibit and stimulate, respectively, glutamate release.^{17,18} The structural basis of this switch is key to understand the heteromer function and the biological advantage behind the GPCR heteromerization phenomenon. Here, we have devised the molecular architecture of the adenosine A₁R-A_{2A}R heteromer in complex with G proteins using a combination of microscope-based single-molecule tracking, molecular modeling, and energy transfer assays in combination with molecular complementation. The results point to A₁ and A_{2A} receptors organizing into a rhombus-shaped heterotetramer that couples to G_i and G_s. The overall structure is very compact and provides interacting interfaces for GPCRs and for G proteins.

3.8.2. Results and Discussion

Reciprocal restriction of adenosine receptor motion in the plasma membrane

To examine the dynamics of A₁R-A_{2A}R heteromers in the plasma membrane of a living cell, the motion of the receptors tagged with fluorescent proteins (A₁R-GFP or A_{2A}R-mCherry) was measured by real time single-particle tracking (SPT) (Figure 1). Examples of fluorescent images and individual receptor trajectories are shown in Additional file 1: Figure S1. Analysis of data corresponding to 500 A₁R-GFP particles shows a linear relationship between the mean-square-displacement (MSD) versus time lag in the trajectories of up to 1,600 single fluorescent particles (Figures 1A, 1C). This is typical for Brownian diffusion, indicating a lack of restrictions in A₁R-GFP motion. Co-expression of A_{2A}R-mCherry (Figure 1B) led to a reduction in lateral mobility of A₁R-GFP, which became confined to plasma membrane regions of $0.461 \pm 0.004 \mu\text{m}$ in diameter. Its diffusion coefficient decreased from $0.381 \pm 0.002 \mu\text{m}^2/\text{s}$ to $0.291 \pm 0.003 \mu\text{m}^2/\text{s}$ ($p=0.002$, one-tailed t-test). Similarly, A₁R-GFP also decreased the A_{2A}R-mCherry diffusion coefficient from $0.317 \pm 0.002 \mu\text{m}^2/\text{s}$ to $0.143 \pm 0.005 \mu\text{m}^2/\text{s}$ ($p < 0.0001$) (Figure

1D-F). A_{2A}R moved within a confinement zone of $0.941 \pm 0.007 \mu\text{m}$ in diameter that was reduced to $0.360 \pm 0.001 \mu\text{m}$ ($p < 0.0001$) when both receptors were co-expressed. We conclude from these mobility comparisons that reciprocally restricted motion of the individual receptor particles must be due to A₁R-A_{2A}R receptor-receptor interactions.

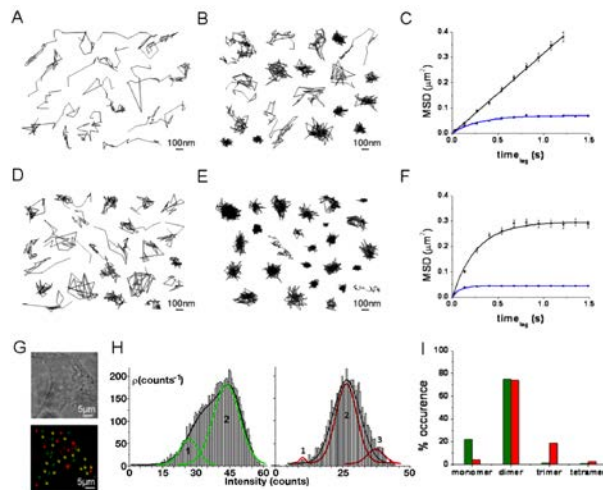


Figure 1 . Cell surface mobility of A₁R-GFP and A_{2A}R-mCherry. Individual trajectories of particles containing enhanced GFP fused to the C-terminus of A₁R (A₁-GFP) (A and B) or mCherry fused to the C-terminus of A_{2A}R (A_{2A}-mCherry) (D and E) on HEK-293T cells expressing A₁-GFP (A), A_{2A}-mCherry (D) or both (B and E). The trajectory and the intensity of the individual fluorescent particles were recorded over time using total internal reflection microscopy (TIRFM) and an electron multiplying (EMCCD) camera recording. Receptor motion was determined by plotting (versus time lag) the mean square displacement (MSD) of A₁-GFP (C) in the absence (*black line*) or presence of A_{2A}-mCherry (*blue line*), or

A_{2A} -mCherry (F) in the presence (*black line*) or presence of A_1 -GFP (*blue line*). Data sets were fitted to mathematical models of free and confined diffusion, for A_1R and $A_{2A}R$ respectively. Panel G. Colocalization of A_1 -GFP and A_{2A} -mCherry is observed (*yellow dots*). Scale bar: 100 nm. Panel H. Distribution of the fluorescence signal of A_1 -eGFP (left) and A_{2A} -mCherry (right) within the stably co-localized receptors (yellow dots). Curves represent approximate amounts of monomers, dimers, or trimers within the co-localized complex. Panel I. Stoichiometry analysis performed for co-localized A_1 -GFP and A_{2A} -mCherry receptor particles co-expressed in HEK-293T cells (yellow dots). Green corresponds to A_1 -GFP and red to A_{2A} -mCherry.

Stoichiometry of A₁ and A_{2A} receptor heterocomplexes

The stoichiometry of the fluorescent receptors on the cell surface can be calculated from the brightness distribution of the individual particles¹⁹ (see Methods). In cells expressing A₁R-GFP, we found the majority of clusters to consist of either two (~47%) or four (~34%) receptors, and clusters with one or three receptors were scarce (~10% and ~9%, respectively) (Additional file 2: Figure S2A and black bars in Additional file 2: Figure S2C). In the case of A_{2A}R-mCherry, the stoichiometry analysis shows that the clusters mostly expressed trimers (45%), being dimers (29%) and tetramers (12%) the second and third most common populations (Additional file 2: Figure S2D and black bars in Additional file 2: Figure S2F). Remarkably, this stoichiometry for either A₁ or A_{2A} receptors was altered when the partner receptor was also expressed. In cells co-expressing A₁R-GFP and A_{2A}R-mCherry the dimer population significantly increased (60% for A₁R-GFP and 50% for A_{2A}R-mCherry, blue bars in Additional file 2: Figures S2C,F) and became the predominant species (Additional file 2: Figures S2B-C,E-F).

In order to focus the analysis on heteromer complexes, we identified individual clusters containing both receptors (individual yellow dots in Figure 1G, displaying both GFP and Cherry

fluorescence). In ~ 1000 analyzed co-localized clusters that consisted of a mixture of A_1 -GFP and A_{2A} -Cherry (yellow dots in Figure 1G), we found a similar high amount of dimers of A_1 R (75%, left panel in Figure 1H and green bar in Figure 1I) and A_{2A} R (74%, right panel in Figure 1H and red bar in Figure 1I). Trimers and tetramers of A_1 R and monomers or tetramers of A_{2A} R were minority or negligible (see Figures 1H-I). In summary, a set percentage of dimers of either A_1 R-GFP or A_{2A} R-mCherry in yellow dots, co-localizing the two receptors, was similar and high ($\sim 75\%$), the heterotetramer containing two A_1 Rs and two A_{2A} Rs was the most predominant species. To our knowledge this is the first stoichiometry data for a GPCR heteromer in living cells.

Arrangement of G proteins interacting with A_1 R and A_{2A} R receptors.

Monomeric GPCRs are capable of activating G proteins.²⁰ However, recent findings suggest that one GPCR homodimer bound to a single G protein may be a common functional unit.²¹ Thus, an emerging question is how G proteins couple to GPCR heteromers. As A_1 R selectively couples to G_i and A_{2A} R to G_s ,²² the working hypothesis was that both G_i and G_s proteins may couple to the A_1 R- A_{2A} R heterotetramers. To test this hypothesis we used

bioluminescence resonance energy transfer (BRET) as says.²³ In agreement with the SPTE experiments (see above), homo- and heterodimers were detected by BRET as says in cells expressing A₁R-Rluc and A₁R-YFP (Figure 2A), A_{2A}R-Rluc and A_{2A}R-YFP (Figure 2B), or A₁R-Rluc and A_{2A}R-YFP (Figure 2E). Neither A₁R-Rluc nor A_{2A}R-YFP interacted with the ghrelin receptor 1a fused to YFP (GHS1a-YFP), used as a control of a protein unable to directly interact with these adenosine receptors (Figure 2A-B). In order to test the presence of the two G proteins in the heterotetramer we transfected cells with minigenes that code for peptides blocking either G_i or G_s binding to GPCRs.²⁴ In addition, cells were also treated with pertussis or cholera toxins that catalyze ADP-ribosylation of G_i or G_s. Clearly, cells treated with pertussis toxin or expressing the minigene-coded peptide that blocks α_i coupling, reduced the value of BRET_{max} for A₁R-A₁R homodimers (Figure 2A) and for A₁R-A_{2A}R heterodimers (Figure 2E) but not for A_{2A}R-A_{2A}R homodimers (Figure 2B). This indicates that G_i is coupled to A₁R in both the homo- and heterodimer. Similarly, blocking G_s-receptor interaction using cholera toxin, or a minigene-coded peptide that blocks α_s coupling, reduced BRET_{max} for A_{2A}R-A_{2A}R

homodimers (Figure 2 B) and for A₁R-A_{2A}R heterodimers (Figure 2E) but not for A₁R-A₁R homodimers (Figure 2A). Interestingly, BRET curves showed sensitivity to both cholera and pertussis toxins in cells expressing either A₁R-Rluc-A₁R-YFP and A_{2A}R (Figure 2 C) or A_{2A}R-Rluc-A_{2A}R-YFP and A₁R (Figure 2 D). Functionality of constructs and controls in cells expressing minigenes, and in cells expressing the ghrelin GHS1a receptor instead of one of the adenosine receptors, are shown in Additional file 3: Figure S 3. To further confirm that G_i binds A_{2A}R in the receptor heteromer, the energy transfer between Rluc fused to the N-terminal domain of the α -subunit of G_i (G_i-Rluc) and A_{2A}R-YFP was analyzed in cells co-expressing or not A₁R (Figure 2 F). A hyperbolic BRET curve was observed only in the presence of A₁R, but not in its absence, indicating that G_i and G_s are bound to their respective receptor homodimers within the A₁R-A_{2A}R heteromer.

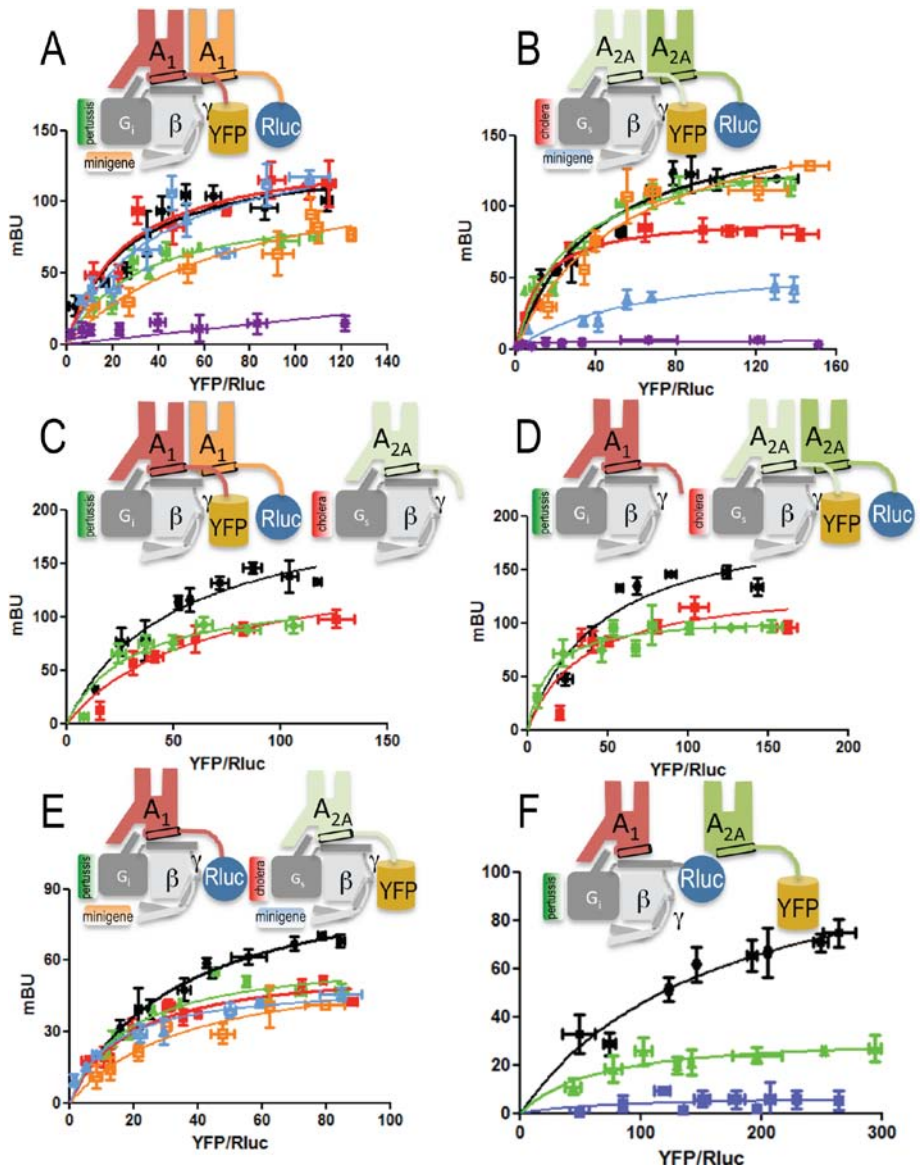


Figure 2 . Influence of G proteins on A₁R and A_{2A}R homo- and heterodimerization. BRET saturation curves were performed in HEK-293T cells 48 h post-transfection with (A, C) 0.3 μ g of cDNA corresponding to A₁R-Rluc and increasing amounts of A₁R-YFP (0.1 to 1.5 μ g cDNA) or GHS1a-YFP (0.25 to 2 μ g cDNA) as negative control (A, purple line), without (A) or with (C) 0.15 μ g of cDNA corresponding to A_{2A}R; (B, D) 0.2 μ g of cDNA corresponding to A_{2A}R-Rluc and increasing amounts of A_{2A}R-YFP (0.1 to 1.0 μ g cDNA) or GHS1a-YFP

(0.25 to 2 μg cDNA) as negative control (B, purple line), without (B) or with (D) 0.5 μg of cDNA corresponding to A_1R ; (E) 0.3 μg of cDNA corresponding to A_1R -Rluc and increasing amounts of A_{2AR} -YFP (0.1 to 1.0 μg cDNA) and (F) 0.5 μg of cDNA corresponding to A_1R (except control blue curves that were obtained in cells not expressing A_1R), 2 μg of cDNA corresponding to G_i -Rluc, and increasing amounts of A_{2AR} -YFP (0.1 to 0.5 μg cDNA). In panels A, B and E, cells were also transfected with 0.5 μg of cDNA corresponding to the G_i - (orange curves) or G_s - (blue curves) related minigenes. Cells were treated for 16h with medium (black curves) or with 10 ng/ml of pertussis toxin (green curves) or 100 ng/ml of cholera toxin (red curves) prior to BRET determination. To confirm similar donor expressions (approximately 100,000 bioluminescence units) while monitoring the increase in acceptor expression (1,000 to 40,000 fluorescence units), fluorescence and luminescence of each sample were measured before energy-transfer data acquisition. MilliBRET unit (mBU) values are the mean \pm SEM of 4 to 6 different experiments grouped as a function of the amount of BRET acceptor. In each panel (top) a cartoon depicts the proteins to which Rluc and YFP were fused and the presence or not of partner receptors and/or G_s or G_i proteins (schemes in C to F are not intended to illustrate stoichiometry as the most predominant form in cells expressing the two receptors is the heterotetramer containing two A_1 and two A_{2A} receptors (see Results)).

Further, two complementary BRET experiments were performed to know the orientation of G_i and G_s within the A_1R - A_{2AR} heterocomplex. First, Rluc and YFP were respectively fused to the N-terminal domains of the α -subunit of G_i (α_i -Rluc) and G_s (α_s -YFP) (Figure 3, a bar) and second, they were fused to the N-terminal domain of the γ -subunit (γ -Rluc and γ -YFP) (Figure 3, b bar). We observed significant energy transfer between γ -Rluc and γ -

YFP in cells co-expressing A₁R and A_{2A}R (Figure 3, b bar) but minimal amounts in negative control cells (Figure 3, c-d bars). In cells expressing either A₁R or A_{2A}R, the energy transfer between γ -Rluc and γ -YFP was also low (Figure 3, e-f bars), suggesting that dimers but not tetramers are the most prevalent form of surface receptors in single-transfected cells. These results in co-transfected cells corroborate the 2:2 stoichiometry obtained from analysis of the fluorescence in single dots and are consistent with G_i and G_s binding to these A₁R-A_{2A}R heterotetramers.

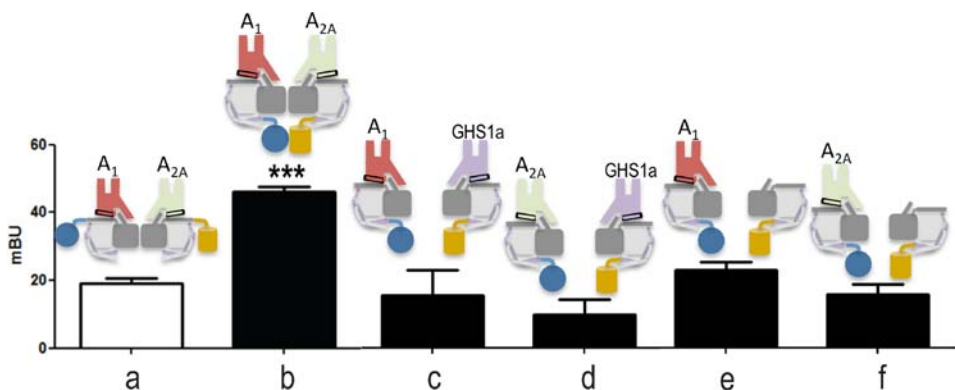


Figure 3. G_s and G_i coupling to a denosine A₁R-A_{2A}R heterocomplexes. BRET experiments were performed in HEK-293T cells 48 h post-transfection with (a, b) 0.2 μ g of cDNA corresponding to A₁R and 0.15 μ g of cDNA corresponding to A_{2A}R, (c, d) 0.2 μ g of cDNA corresponding to A₁R or 0.15 μ g of cDNA corresponding to A_{2A}R and 0.4 μ g of cDNA corresponding to growth hormone secretagogue receptor GHS1a, (e) 0.2 μ g of cDNA corresponding to A₁R, or (f) 0.15 μ g of cDNA corresponding to A_{2A}R. Cells were also transfected

with 2 μg of cDNA corresponding to α -subunit of G_i fused to Rluc and increasing amounts of cDNA corresponding to α -subunit of G_s fused to YFP (a) or 0.3 μg of cDNA corresponding to γ -subunit fused to Rluc and increasing amounts of cDNA corresponding to γ -subunit fused to YFP (b-f). Maximum miliBRET unit (mBU) values are the mean \pm SEM of 4 different experiments. A scheme showing the protein to which Rluc and YFP were fused is provided (top). *** $p < 0.001$ by one-way ANOVA with post hoc Dunnett's test.

Molecular model of G_i and G_s bound to the A_{1R} - A_{2AR} heterotetramer

To identify the orientation of the G protein in the receptor homodimer, we combined energy-transfer assays between α_s -Rluc (Rluc at the N-terminus of the G protein α -subunit) and A_{2AR} -YFP (Figure 4A) with transmembrane (TM) interfaces based on crystal structures of GPCRs,^{3,4} which have been recently summarized.²⁵ The observed high-energy transfer using α_s - and A_{2AR} -YFP indicated a close proximity between the N-tail of the α -subunit of G_s and the C-tail of A_{2AR} . Interestingly, Rluc and YFP in the 'monomeric' A_{2AR} - G_s complex (see Methods), point toward distant positions in space (Figure 4B). Therefore, the observed BRET should occur between Rluc in the G protein α -subunit and a second A_{2AR} -YFP protomer. Among all described TM interfaces for receptor homodimerization (see Additional file 4: Figure S4), we

propose the TM4/5 interface, which is observed in the oligomeric structure of β_1 -AR⁴ and with structures derived from coarse-grained molecular dynamic simulations.²⁶ In fact, this is the only interface that favors BRET between α_s -Rluc and a second A_{2A}R-YFP protomer in a homodimer (Figure 4C). The homologous A₁R homodimer was built using the same TM4/5 interface as for A_{2A}R (see Additional file 4: Figure S4 and its legend).

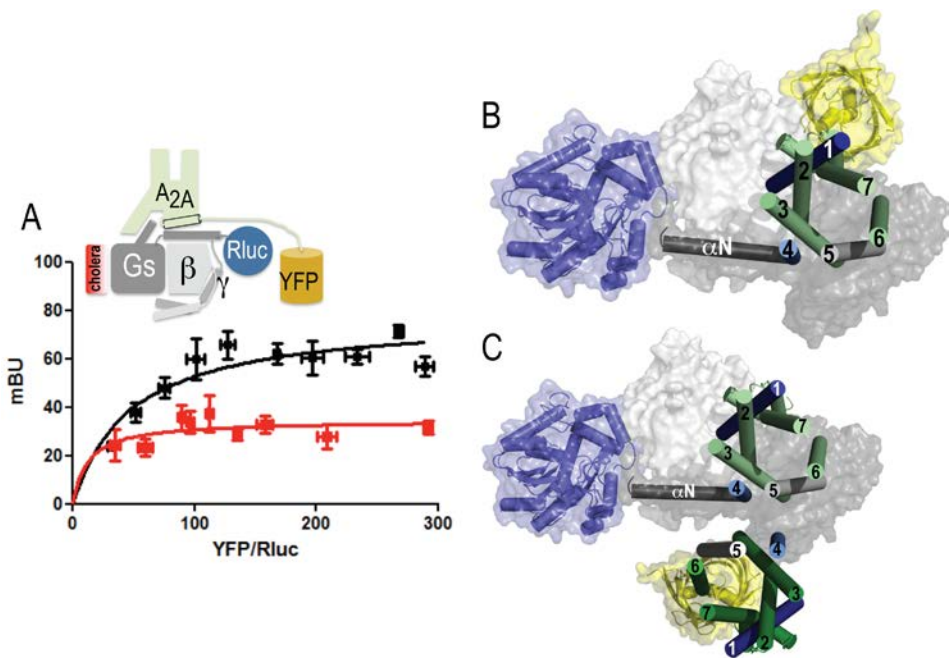


Figure 4 . Orientation of a G protein in a receptor homodimer. BRET saturation experiments were performed in HEK-293T cells transfected with 2 μ g of cDNA corresponding to α -subunit of G_s fused to Rluc and increasing amounts of A_{2A}R-YFP (0.1 to 0.5 μ g) cDNA. Panel A. BRET measurements in cells pre-

treated for 16h with medium (black line) or with 100 ng/ml of cholera toxin (red line). Both fluorescence and luminescence of each sample were measured before every experiment to confirm similar donor expressions (approximately 50,000 bioluminescence units) while monitoring the increase in acceptor expression (1,000 to 10,000 fluorescence units). miliBRET unit (mBU) values are the mean \pm SEM of 4-5 different experiments grouped as a function of the amount of BRET acceptor. A scheme of the placement of donor and acceptor BRET moieties is provided (top). Panel B. Molecular model of the A_{2A}R-G_s complex. Rluc (in blue) is attached to the N-terminal α N helix of G_s (in gray) and YFP (in yellow) is attached to the C-terminal domain of A_{2A}R (in light green) (see Additional file 9 : Figure S 9 for details). Panel C. Arrangement of A_{2A}R homodimers modeled via the TM4/5 interface as observed in the oligomeric structure of β_1 -AR⁴. The A_{2A}R protomer bound to α_s is shown in light green, whereas the second A_{2A}R-YFP protomer is shown in dark green. This molecular model in panel C (center-to-center distance between Rluc and YFP of 6.5 nm), in contrast to the model shown in panel B (8.3 nm), would favor the observed high-energy transfer (see panel A) between α_s -Rluc and A_{2A}R-YFP.

The remaining possible TMs able to form heteromeric interfaces are TM1 and TM5/6 (Figure 5). Both are possible inter-GPCR interfaces as observed in the structure of the μ -opioid receptor (μ -OR).³ To discern between these two possibilities a bimolecular fluorescence complementation strategy was undertaken. For such purpose the N-terminal fragment of Rluc8 was fused to A₁R (A₁R-nRluc8) and its C-terminal domain to A_{2A}R (A_{2A}R-cRluc8), which only upon complementation can act as a BRET donor (Rluc8). The

BRET acceptor protein was obtained upon complementation of the N-terminal fragment of YFP Venus protein fused to A₁R (A₁R-nVenus) and its C-terminal domain fused to A_{2A}R (A_{2A}R-cVenus). When all four receptor constructs were transfected we obtained a positive and saturable BRET signal (BRET_{max} of 35 ± 2 mBU and BRET₅₀ of 16 ± 3) that was not obtained for negative controls (Additional file 5: Figure S5). Figure 5A-B shows that the hemi-donor (A₁R-nRluc8 and A_{2A}R-cRluc8) and the hemi-acceptor (A₁R-nVenus and A_{2A}R-cVenus) moieties, placed at the C-terminus of the receptors, can only complement if A₁R-A_{2A}R heterodimerization occurs via the TM5/6 interface. TM4/5 for homo- and TM5/6 interface for hetero-dimerization give a rhombus-shaped tetramer organization (Figure 5A). Remarkably, cell pre-incubation with either pertussis or cholera toxins decreased the BRET_{max} by 35% (Figure 5C), further suggesting that both G_s and G_i proteins bind to the A₁R-A_{2A}R heterotetramer.

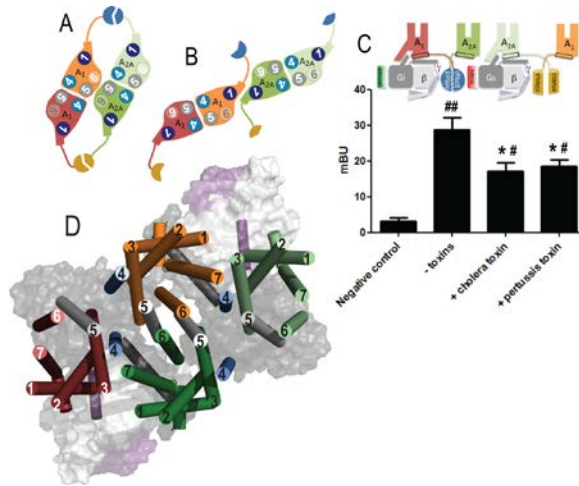


Figure 5. BRET-aided construction of a model consisting of G_i and G_s bound to an A₁R-A_{2A}R heterotetramer. Panels A-B. A₁R-A_{2A}R receptor tetramer built using TM5/6 (A) or TM1 (B) inter-receptor interfaces modeled as in the structure of the μ -opioid receptor³. TM helices involved in receptor dimerization: 1, 4, and 5, are highlighted in dark blue, light blue, and gray, respectively. nRluc8 and cRluc8 are shown in blue and nVenus and cVenus in dark yellow. Panel C: BRET and bimolecular fluorescence complementation experiments were performed in HEK-293T cells transfected with 1.5 μ g of cDNA corresponding to A₁R-cRluc8 and A_{2A}R-nRluc8 and 1.5 μ g of cDNA corresponding to A₁R-nVenus and A_{2A}R-cVenus. Negative control corresponds to cells transfected with 1 μ g of cDNA corresponding to nRluc8 and 1.5 μ g of cDNA corresponding to A_{2A}R-nRluc8, A₁R-nVenus and A_{2A}R-cVenus. Cells were treated for 16 h with medium (-toxins) or with 10 ng/ml of pertussis toxin (+pertussis) or 100 ng/ml of cholera toxin (+cholera) prior to BRET determination. The relative amount of BRET is given as in Figure 4 and values are mean \pm S.E.M. of 3 different experiments. Student's t test showed statistically significant differences respect to control (#p<0.05, ##p<0.01) and a significant effect in the presence of either toxin over BRET in the absence of toxins (*p<0.05). At top, a schematic representation showing the protein to which the hemiluminescent or fluorescent proteins were fused. Panel D. Molecular model of the A₁R-A_{2A}R tetramer in complex with G_i and G_s. The color code of the depicted complex is: A₁R bound to G_i is shown in

red, G_i -unbound A_{1R} in orange, $A_{2A}R$ bound to G_s in dark green, G_s -unbound $A_{2A}R$ in light green, and the α , β -, and γ -subunits of G_i and G_s in dark grey, light grey and purple, respectively. TM helices 4 and 5 are highlighted in light blue and gray, respectively.

We next evaluated, by computational tools, whether the proposed A_{1R} - $A_{2A}R$ heterotetramer could couple to both G_i and G_s proteins. Clearly, the external protomers of the proposed A_{1R} - $A_{2A}R$ heterotetramer can bind G_i and G_s proteins (Figure 5D). This model positions the α -subunits of G_i and G_s in close contact facing the interior of the tetrameric complex, while the N-terminal α -helices of α_i - and α_s - point outside the complex. The N-terminal α -helices of the γ -subunits are in close proximity facing to the inside (Additional file 6: Figure S 6), which explains the significant energy transfer observed between γ -Rluc and γ -YFP (Figure 3, b bars). The model provides experimental insights into the structural arrangement of heteromers consisting of two GPCRs and coupled to two G proteins, the possibility of which has been recently discussed.²⁵ We used molecular dynamics simulations (MD) to study the stability of this complex. Additional file 7: Figure S7 shows root-mean-square deviations (rmsd) on protein α -carbons throughout the MD

simulation, as well as key intermolecular distances among protomers and G proteins. Clearly, A₁R and A_{2A}R, both bound and unbound to the G protein, G_i and G_s, maintain a close structural similarity (rmsd \approx 0.3 nm) relative to the initial structures (Additional file 7: Figure S7A). The fact that rmsd values of the whole system, formed by the A₁R-A_{2A}R heterotetramer bound to G_i and G_s, are of the order of 0.6 nm suggests that the initial structural model is maintained during the MD simulation (Additional file 7: Figure S7A). As a consequence, selected intermolecular distances among protomers and G proteins remain constant during the MD simulation (Additional file 7: Figure S7B). A key process in the assembly of the heterotetramer is the TM interfaces for homo- (TM4/5) and hetero- (TM5/6) dimerization. Additional file 8: Figure S8B shows rmsd values of the four-helix bundle forming the TM4/5 and TM5/6 interfaces, the initial and final snapshots of these bundles, as well as the evolution of the A₁R-A_{2A}R heterotetramer during the MD simulation. Clearly, the rather small structural variations of these four-helix bundle, also reflected in rmsd < 0.3 nm, suggest a stable complex. Notably, the TM5/6 four-helix bundle seems more stable than the TM4/5 bundles as shown by its lower rmsd values. Additional file 8: Figures S8B and C depict

contact maps of the TM4/5 and TM5/6 interfaces, as well as the evolution of the network of hydrophobic interactions within these interfaces during the MD simulation.

3.8.3. Conclusions

For more than a decade, experimental evidence has supported the occurrence of homo- and hetero-oligomers of GPCRs.²¹ However, our basic understanding of what makes heteromers different than homomers remains unknown. Our results, studying adenosine receptors as a model heteromer, point to three important new findings. First, the predominant stoichiometry in cells expressing A₁R-A_{2A}R heteromers is 2:2, i. e. dimer of dimers (tetramer). Second, two different heterotrimeric G proteins can couple to heteromers, the overall complex constituting a functional unit. Third, the molecular orientation within the heteromer complex affords various qualitatively different interfaces; the two more relevant are: the inter-protomer heteromeric interface and the inter-G-protein interface. Presumably, the two interfaces provide the key characteristic of heteromers: the ability of one protomer/G protein complex to influence the signaling of the other. Surely, allosteric effects occurring between heteroreceptors and between G_s and G_i

proteins are due to conformational changes transmitted along the intimately interacting molecules in the complex. The fact that, in our controlled cell transfection system expressing low density of receptors, minor species formed by monomers and trimers were found in addition to a high predominance of tetramers in the plasma membrane strongly supports the occurrence of an *in vivo* dynamic distribution of receptors.

Adenosine was, from an evolution point of view, one of the first extracellular regulators as it is involved in energy and nucleic acid metabolisms. Adenosine A_1 and A_{2A} receptors are expressed in almost every mammalian organ/tissue. In the heart, where adenosine plays a key role in both inotropic and chronotropic regulation, A_1R -mediated cardioprotection did not occur in $A_{2A}R$ knockout mice, suggesting an interaction between A_1 and A_{2A} receptors. In neurons, A_1 and A_{2A} receptors show colocalization, thus, leading to inter-receptor interactions unveiled by pharmacological treatments. For instance, Okada et al.²⁷ showed that cAMP-dependent protein kinase A plays a role in the regulation of hippocampal serotonin release mediated by both A_1 and A_{2A} receptors. Similarly, the control of γ -amino butyric acid transport in astrocytes was attributed

to the expression of A_1R - $A_{2A}R$ heteromers and to a specific mechanism by which the heteromer signals via G_i or via G_s depending on the concentration of adenosine.²⁸ The structural basis of the differential signaling by the heteromer-G protein macromolecular complex likely implies communication at the receptor-receptor level but also between G_s and G_i . Because the binding of two G proteins to a heterodimer is not feasible due to steric clashes, our finding that the A_1R - $A_{2A}R$ heterotetramer may bind to both G_s and G_i provides a structural framework to interpret experimental data.

3.8.4. Methods

Total internal reflection single-molecule microscopy and single particle data analysis

Single-molecule imaging and tracking were performed on a Nikon Total Internal Reflection Fluorescence (TIRF) system as detailed in Supplementary Methods. Typically 500 readouts of a 512 x 512 pixels region, the full array of the CCD chip were acquired. For single particle data analysis parameters were calculated applying the equations described in Supplementary Methods.

Cell culture and transient transfection

CHO and HEK-293T cells were grown at 37°C in Dulbecco's modified Eagle's medium (DMEM) (Gibco) supplemented with 2mM L-glutamine, 100 U/ml penicillin/streptomycin, and 5% (v/v) heat inactivated Foetal Bovine Serum (FBS) (all supplements were from Invitrogen, Paisley, Scotland, UK). Cells were transiently transfected with cDNA corresponding to receptors, fusion proteins, A_{2A}R mutants or G protein minigene vectors obtained as detailed in Expanded View by the PEI (PolyEthylenImine, Sigma) method. To control the cell number, sample protein concentration was determined using a Bradford assay kit (Bio-Rad, Munich, Germany) using bovine serum albumin dilutions as standards. For single-molecule imaging, cells were seeded into six-well plates containing glass coverslips (No. 1, round, 24 mm; Assistent, Sondheim, Germany) or into Lab-Tek Chambered #1.0 Borosilicate Coverglass System (Nunc, Thermo Fisher Scientific, Schwerte, Germany). Cell transient transfections were performed with LipofectamineTM2000 (Invitrogen, Life Technologies, Darmstadt, Germany) or FuGENE 6 (Roche Applied Science, Indianapolis, IN, USA) and application of

0.1-0.2 μ g plasmid DNA per well. Before each experiment cells were washed three times with 200 μ L phenol red-free DMEM.

Plasmids

Sequences encoding amino acid residues 1-155 and 155-238 of YFP Venus protein, and amino acids residues 1-229 and 230-311 of RLuc8 protein were subcloned in the pcDNA3.1 vector to obtain the YFP Venus and RLuc8 hemi-truncated proteins. The human cDNAs for adenosine receptors $A_{2A}R$ and A_1R , cloned into pcDNA3.1, were amplified without their stop codons using sense and antisense primers harboring unique EcoRI and BamHI sites to clone receptors in pcDNA3.1RLuc vector (pRLuc-N1 PerkinElmer, Wellesley, MA) and EcoRI and KpnI to clone $A_{2A}R$, A_1R or ghrelin 1a receptor, GHS1a, in pEYFP-N1 vector (enhanced yellow variant of GFP; Clontech, Heidelberg, Germany). $G_{\alpha s}$ cloned in *SFVI* vector, $G_{\alpha i}$ cloned in the pcDNA3.1 vector or G_{γ} cloned in *pEYFP-C1* vector, were amplified without their stop codons using sense and antisense primers harboring unique *HindIII* and *BamHI* sites to clone them into the pcDNA3.1-Rluc vector or EcoRI and KpnI to clone $G_{\alpha s}$ into the pEYFP-N1 vector. The amplified fragments were subcloned to be in-frame with restriction sites of pcDNA3.1RLuc or

pEYFP-N1 vectors to give the plasmids that express proteins fused to RLuc or YFP on the N-terminal ($G_{\alpha s}$ -RLuc, $G_{\alpha i}$ -RLuc, G_{γ} -RLuc, $G_{\alpha s}$ -YFP and G_{γ} -YFP) or on the C-terminal end (A_1R -RLuc, $A_{2A}R$ -RLuc, A_1R -YFP, $A_{2A}R$ -YFP and GHS1a-YFP). The human cDNA for A_1R or GHS1a were subcloned into pcDNA3.1-nRLuc8 or pcDNA3.1-nVenus to give plasmids that express A_1R or GHS1a fused to either nRLuc8 or n-YFP Venus on the C-terminal end of the receptor (A_1R -nRLuc8 and A_1R -nVenus or GHS1a-nRLuc8 and GHS1a-nVenus). The human cDNA for $A_{2A}R$ or GHS1a were subcloned into pcDNA3.1-cRLuc8 or pcDNA3.1-cVenus to give plasmids that express receptors fused to either cRLuc8 or cYFP Venus on the C-terminal end of the receptor ($A_{2A}R$ -cRLuc8 and $A_{2A}R$ -cVenus or GHS1a-cRLuc8 and GHS1a-cVenus). Expression of constructs was tested by confocal microscopy and the receptor-fusion protein functionality by second messengers, ERK1/2 phosphorylation and cAMP production as described previously.^{13,14,17,29}

“Minigene” plasmid vectors are constructs designed to express relatively short polypeptide sequences following their transfection into mammalian cells. Here we used minigene constructs encoding

the carboxyl-terminal 11 amino acid residues from G_{α} subunits of $G_{i1/2}$ (G_i minigene) or G_s (G_s minigene) G proteins that inhibit G-protein coupling to the receptor and consequently inhibit the receptor-mediated cellular responses as previously described.²⁴ The cDNA encoding the last 11 amino acids of human G_{α} subunit corresponding to $G_{i1/2}$ (I K N N L K D C G L F) or G_s (Q R M H L R Q Y E L L), inserted in a pcDNA 3.1 plasmid vector were generously provided by Dr. Heidi Hamm.

Energy Transfer Assays

For BRET and complementation BRET assays, HEK-293T cells were transiently cotransfected with a constant amount of cDNA encoding for proteins fused to RLuc, nRLuc8 or cRLuc8, and with increasing amounts of the cDNA corresponding to proteins fused to YFP, nYFP Venus or cYFP Venus (see Figure legends). To quantify protein-YFP expression or protein-reconstituted YFP Venus expression, cells (20 μ g protein) were distributed in 96-well microplates (black plates with a transparent bottom) and fluorescence was read in a FluoStar Optima fluorimeter (BMG Labtechnologies, Offenburger, Germany) equipped with a high-energy xenon flash lamp, using a 10 nm bandwidth excitation filter

at 400 nm reading. Protein fluorescence expression was determined as fluorescence of the sample minus the fluorescence of cells expressing the BRET donor alone. For BRET measurements, the equivalent of 20 μ g of cell suspension were distributed in 96-well microplates (Corning 3600, white plates; Sigma) and 5 μ M coelenterazine H (Molecular Probes, Eugene, OR) was added. After 1 minute for BRET or after 5 minutes for BRET with bimolecular fluorescence complementation, the readings were collected using a Mithras LB 940 that allows the integration of the signals detected in the short-wavelength filter at 485 nm (440-500 nm) and the long-wavelength filter at 530 nm (510-590 nm). To quantify protein-RLuc or protein-reconstituted RLuc8 expression luminescence readings were also performed after 10 minutes of adding 5 μ M coelenterazine H. The net BRET is defined as $[(\text{long-wavelength emission}) / (\text{short-wavelength emission})] - C_f$ where C_f corresponds to $[(\text{long-wavelength emission}) / (\text{short-wavelength emission})]$ for the donor construct expressed alone in the same experiment. BRET is expressed as miliBRET units, mBU (net BRET x 1,000).

Computational model of the A₁R-A_{2A}R tetramer in complex with G_i and G_s

The crystal structure of inactive A_{2A}R (PDB code 4EIY)³⁰ was used for the construction of human A_{2A}R (Uniprot entry P29274) and A₁R (P30542) homology models using Modeller 9.12.³¹ These receptors share 51% of sequence identity and 62% of sequence similarity, excluding the C-term after helix 8. ICL3 of A_{2A}R (Lys209-Gly218) and A₁R (Asn212-Ser219) were modeled using Modeller 9.12³¹ using ICL3 of squid rhodopsin (PDB code 2Z73) as a template. The C-terminus tails of A₁R, containing 16 amino acids (Pro311-Asp326), and of A_{2A}R, containing 102 amino acids (Gln311-Ser412), were modeled as suggested for the OXE receptor³² (see Additional file 9: Figure S9 for details). The N-terminus of A₁R and A_{2A}R were not included in the model. The “active” conformations of A₁R bound to G_i and A_{2A}R bound to G_s were modeled from the crystal structure of β₂-AR in complex with G_s (PDB code 3SN6).³³ The globular α-helical domain of the α-subunit was modeled in the “closed” conformation,³⁴ using the crystal structure of [AlF₄⁻]-activated G_i (PDB code 1AGR). The location of YFP (PDB code 2RH7) attached to the C-tail of A_{2A}R

was determined as suggested for the OXE receptor³² (see Additional file 9: Figure S9 for details). Rluc (PDB code 2PSD) and YFP were fused to the to the N-terminus of the α - and γ - subunits of G_i and G_s by a covalent bond. The structures of adenosine receptor oligomers were modeled via the TM4/5 interface for homo-dimerization, using the oligomeric structure of the β_1 -AR (PDB code 4GPO),⁴ or via the TM5/6 interface for hetero-dimerization, using the structure of the μ -OR (PDB code 4DKL).³ The G_i -bound A_1R and G_s -bound $A_{2A}R$ protomers were rotated 10° to avoid the steric clash of the N-terminal helix of G_i and G_s with the C-terminal helix (Hx8) of G_s -unbound $A_{2A}R$ and G_i -unbound A_1R , respectively. This computational model, without Rluc and YFP, was placed in a rectangular box containing a lipid bilayer (814 molecules of POPC) with explicit solvent (102,973 water molecules) and a 0.15 M concentration of Na^+ and Cl^- (1,762 ions). This initial complex was energy-minimized and subsequently subjected to a 10 ns MD equilibration, with positional restraints on protein coordinates. These restraints were released and 500 ns of MD trajectory were produced at constant pressure and temperature (see Additional file 10: Movie M1). Computer simulations were performed with the

GROMACS 4.6.3 simulation package,³⁵ using the AMBER99SB force field as implemented in GROMACS and Berger parameters for POPC lipids. This procedure has been previously validated.³⁶

3.8.5. Supporting information

Single particle data analysis

Data processing was performed using Matlab (MathWorks, Natick, MA, USA). By correlation analysis between consecutive images the two dimensional trajectories of individual molecules in the plane of focus were reconstructed by determining the probability and setting a high-confidence threshold that each step in a trajectory was from the same particle. Multiple data sets were produced for every receptor type and for the existing complexes of the receptors separately. In brief, trajectories were then analyzed as described previously.¹ For the analysis of the (r_i^2, t_{lag}) plots, a positional accuracy of 14 ± 3 nm was considered in our measurements.²

The lateral diffusion of Brownian particles in a medium characterized by a diffusion constant D is described by the

cumulative probability distribution function for the square displacements, r :^{2,3}

$$\tilde{P}(r^2, t_{lag}) = 1 - \exp\left(-\frac{r^2}{r_0^2}\right) \quad [\text{Eq. 1}]$$

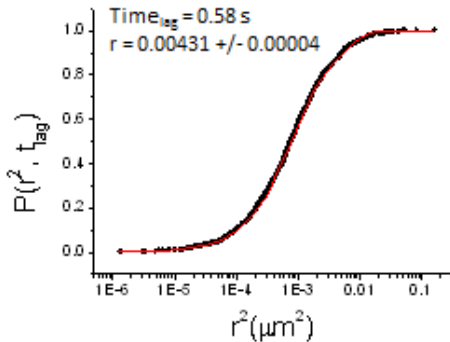
$P(r^2, t_{lag})$ is the probability that the Brownian particle starting at the origin will be found within a circle of radius r at time t_{lag} . Provided that the system under study segregates into two components, characterized by mean-square displacements r_1^2 and r_2^2 , and relative fractions α and $(1 - \alpha)$, respectively, equation 1 becomes:^{1,3}

$$\tilde{P}(r^2, t_{lag}) = 1 - \left[\alpha \cdot \exp\left(-\frac{r^2}{r_1^2}\right) + (1 - \alpha) \cdot \exp\left(-\frac{r^2}{r_2^2}\right) \right] \quad [\text{Eq. 2}]$$

The cumulative probability distributions $P(r_i^2, t_{lag})$ were constructed for each time lag from the single-molecule trajectories by counting the number of square displacements with values $< r^2$, and subsequent normalization by the total number of data points³. Probability distributions with $n > 1,000$ data points were least-square fitted to equation 2, resulting in a parameter set $\{r_1^2(t_{lag}), r_2^2(t_{lag}), \alpha\}$, for each time lag, t_{lag} . This approach of fitting leads to a robust

estimation of the mean-square displacements r_i^2 even when the mobility is not purely random.¹

An example of the data is provided in Graph 1.



Graph 1. Example of the mobility probability distribution. Example from data using cells expressing A₁R- eGFP at $t_{lag} = 0.58$ s

For mobility analysis, the diffusional behavior of the respective populations of molecules was revealed by plotting the mean square displacement (r_i^2) versus t_{lag} . The (r_i^2 , t_{lag}) data sets were fitted by a free diffusion model,

$$r_i^2(t_{lag}) = 4D_i t_{lag} \quad [\text{Eq. 3}]$$

where r_i^2 is proportional to time t_{lag} . When diffusion is hindered by obstruction or trapping in such a way that the mean square displacement is proportional to some power of time < 1 ($r_i^2 \sim t^\alpha$, $\alpha < 1$) (anomalous subdiffusion), the diffusion constant becomes:¹

$$D = \Gamma t_{lag}^{1-\alpha}$$

[Eq. 4]

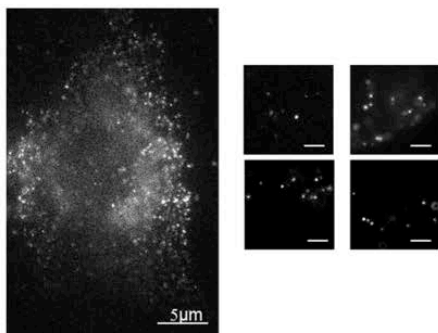
If $\alpha = 1$, then $r_i^2 \sim 1$, $D = \Gamma$ is constant and diffusion is normal. The confined diffusion model assumes that diffusion is free within a square of side length L , surrounded by an impermeable, reflecting barrier. Then the mean-square displacement depends on L and the initial diffusion coefficient D_0 , and varies with t_{lag} as:^{1,4}

$$r_i^2(t_{lag}) = \frac{L^2}{3} \left[1 - \exp\left(\frac{-12D_0 t_{lag}}{L^2}\right) \right]$$

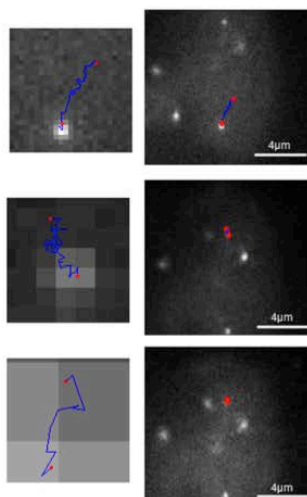
[Eq. 5]

Fluorescence distribution may be used to determine local stoichiometry.^{5,6} The only difference between fluorescence of small fluorophore clusters and single fluorophores is the higher intensity. Fluorophore photobleaching or blinking has a significant impact on an average intensity of a fluorophore cluster, reducing it and this way making a direct fluorescence intensity count more complicated. The probability density function of the fluorescence intensity displays a discrete structure and may be fitted with multiple Gaussians models to calculate the molecular stoichiometry^{5,6}.

A



B



C

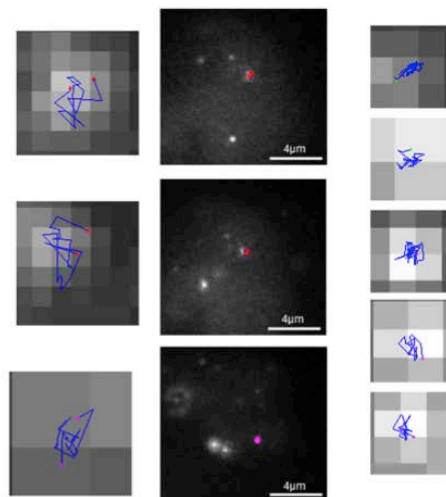


Figure S1. Examples of receptor trajectories in HEK-293T cells.

Images of cells expressing A₁R-eGFP (A) and of particular trajectories of A₁R-eGFP- (B) or A_{2A}R-mCherry- (C) containing particles.

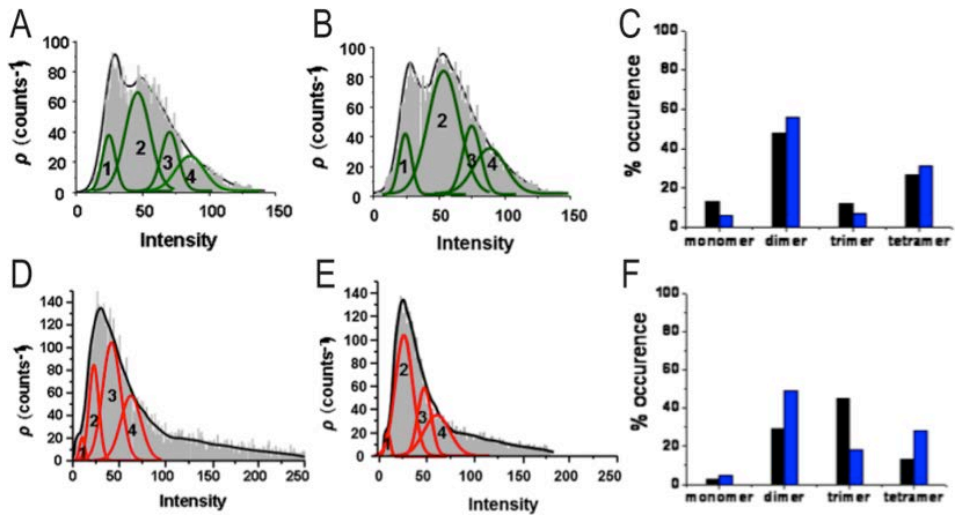


Figure S2 . Graphical d escription o f t he s toichiometry of A₁R-GFP, A_{2A}R-mCherry or both A₁-eGFP and A_{2A}-mCherry.

The fluorescence intensity signal distribution (*grey area*) detected for more than 7000 independent observations is given for HEK-293T cells expressing A₁-eGFP (A), A_{2A}-mCherry (D), or both A₁-eGFP and A_{2A}-mCherry (B, E). The stoichiometry analysis was performed for A₁-eGFP (A-B) and A_{2A}-mCherry (D-E). Curves showing approximate amounts of monomers, dimers, trimers and tetramers were also displayed in green for A₁-eGFP (A-B) and in red for A_{2A}-mCherry (D-E). The occurrence of monomers, dimers, trimers and tetramers for A₁-eGFP (C) expressed alone (*black bars*) or in the presence of A_{2A}-mCherry (*blue bars*) or A_{2A}-mCherry (F) alone (*black bars*) or in the presence of A₁-eGFP (*blue bars*) on the cell surface was calculated by stoichiometry analysis from results shown in (A-B, D-E).

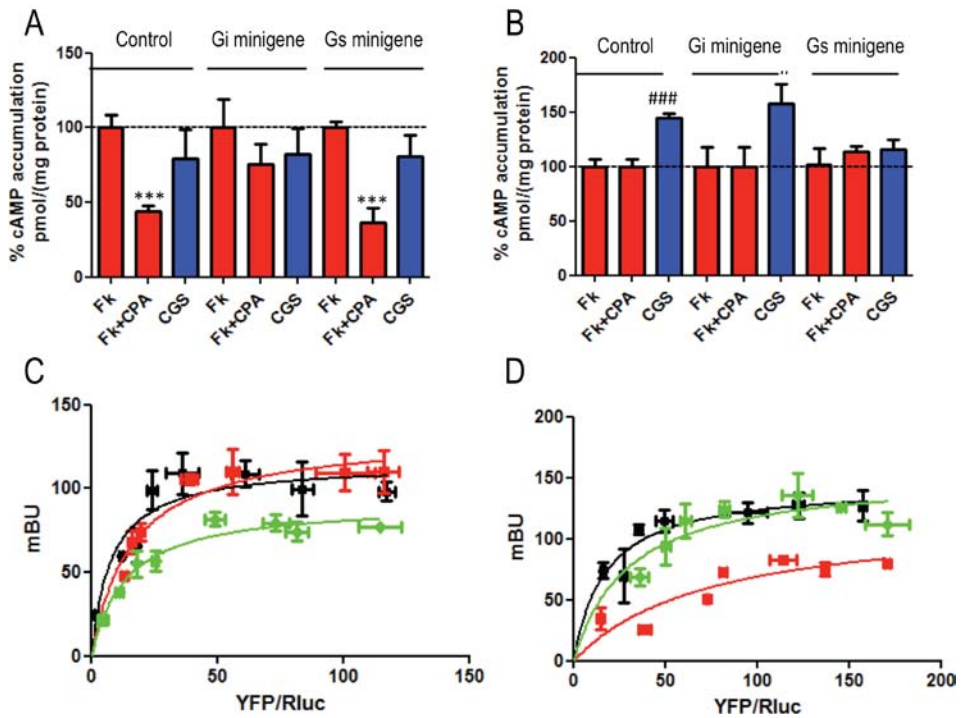


Figure S 3. C ontrols of c AMP p roduction an d B RET as says i n c ells expressing minigenes a nd i n cells ex pressing t he ghrelin GHS1a recep tor instead of one of the adenosine receptors.

Panels A-B. cAMP determination in HEK-293T cells transfected with 0.3 μ g of cDNA corresponding to A₁R (A) or with 0.2 μ g of cDNA corresponding to A_{2A}R (B) plus (control)/minus 0.5 μ g of cDNA corresponding to minigenes coding for peptides blocking either Gi or Gs binding. Cells were stimulated with the A₁R agonist CPA (10 nM, red bars) in the presence of 0.5 μ M forskolin. or with the A_{2A}R agonist CGS 21680 (200 nM, blue bars). Values expressed as % of the forskolin treated cells (red bars) or of the basal (blue bars) are given as mean \pm SD (n=4-8). One-way ANOVA followed by a Bonferroni *post hoc* test showed a significant effect over forskolin (red bars, *** p < 0.001) or over basal (blue bars, ## p < 0.01, ### p < 0.001).

Panels C -D. BRET saturation curves were performed in HEK-293T cells transfected with (C) 0.3 μ g cDNA coding for A₁R-Rluc, increasing amounts of

cDNA coding for A₁R-YFP (0.1 to 1.5 μg cDNA) and 0.4 μg cDNA coding for GHS1a, or (D) with 0.2 μg of cDNA coding for A_{2A}R-Rluc, increasing amounts of cDNA coding for A_{2A}R-YFP (0.1 to 1.0 μg cDNA) and 0.5 μg cDNA coding for t o G HS1a. P rior t o B RET de termination c ells were t reated f or 16h with medium (black curves), with 10 ng/ml of p ertussis toxin (green curves) or with 100 ng/ml of cholera toxin (red curves). mili BRET units (mBU) are given as the mean ± SD (n=4-6 different experiments grouped as a function of the amount of BRET acceptor).

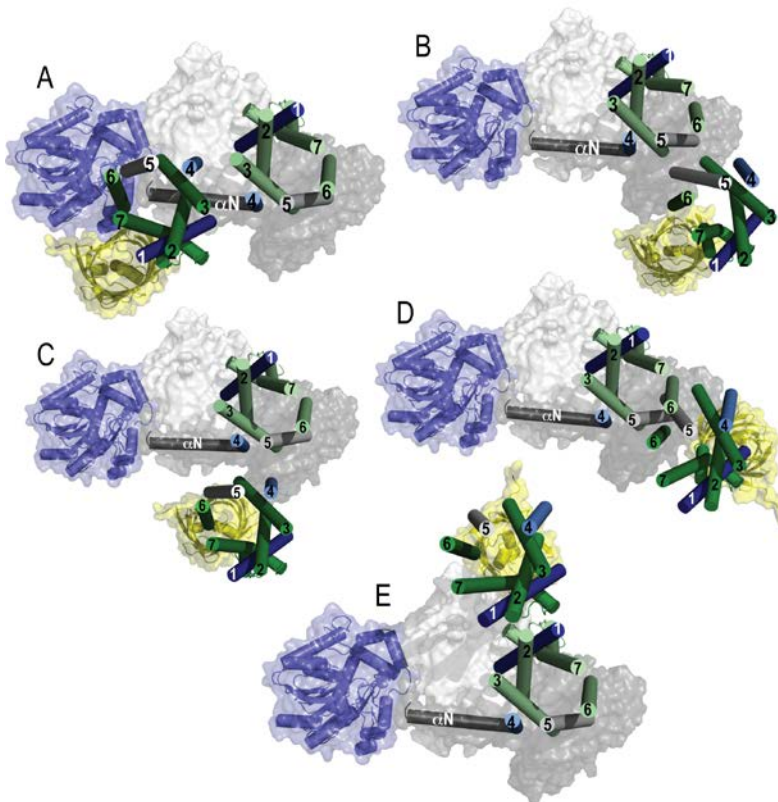


Figure S4. Possible interfaces in A_{2A}R homodimers in complex with G_s

In (A-E) the A_{2A}R homodimer was modeled through TM4 using the H₁- receptor structure as a template (panel A), through TM5 using the structure of s quid

rhodopsin (panel B), through TM4/5 using the β_1 -receptor structure (panel C), and via TM5/6 (panel D) and TM1 (panel E) using the μ -opioid receptor structure. TM helices 1, 4, and 5 involved in receptor dimerization are highlighted in dark blue, light blue, and gray, respectively. $A_{2A}R$ protomers bound to G_s (in gray) are shown in light green, whereas G_s -unbound $A_{2A}R$ protomers are shown in dark green. Rluc (in blue) is attached to the N-terminal αN helix of G_s and YFP (in yellow) is attached to the C-terminal domain of the G_s -unbound $A_{2A}R$ protomer (in light green). It is important to note that the position of YFP is highly dependent on the orientation of the long and highly flexible C-tail of $A_{2A}R$ (102 amino acids, Gln311-Ser412), which was modeled as described for the OXE receptor (OXER)³² (see Additional file 9: Figure S9 for details). Despite these limitations, we can crudely estimate the approximate distances between the center of mass of Rluc and YFP: 4.6, 10.1, 6.5, 11.6, 8.3 nm for panels A-E, respectively. Thus, among all these possible dimeric interfaces, only molecular models depicted in panels A (TM4 interface) and C (TM4/5 interface) would favor the observed high-energy transfer between G_s -Rluc and $A_{2A}R$ -YFP (Figure 4A in main paper). However, there is a steric clash between the N-terminal helix of G_s and the dark-green protomer in the TM4 interface. Accordingly, we have modeled $A_{2A}R$ homodimerization via the TM4/5 interface. Unfortunately, similar experiments with cells transfected with G_i -Rluc and A_1R -YFP could not be accomplished because of lack of receptor expression (not shown); probably the shorter C-tail of A_1R (16 amino acids, Pro311-Asp326) could not accommodate YFP in the presence of G_i in the right three-dimensional structure. The A_1R homodimer was built using the same TM4/5 interface as for $A_{2A}R$.

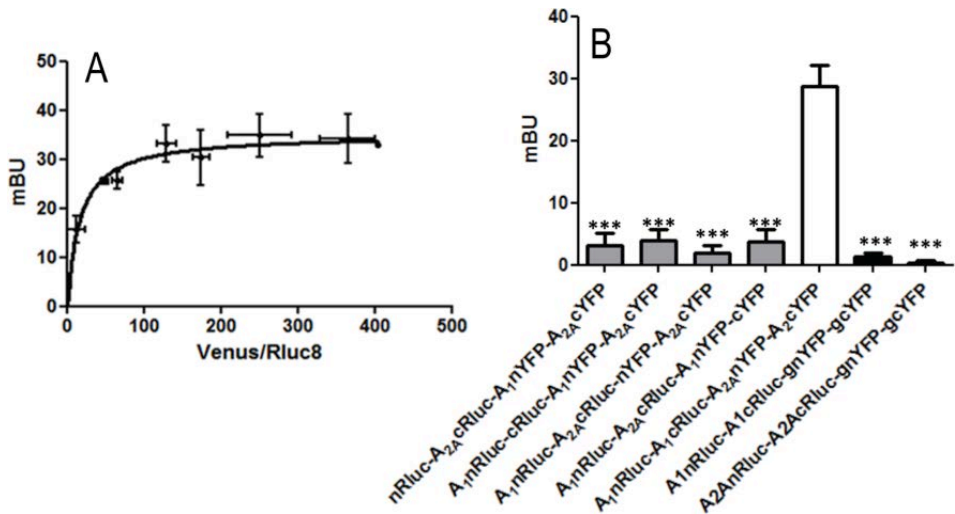


Figure S5. BRET assays in cells expressing fusion proteins containing hemi Rluc8 and hemi Venus moieties and expressing the ghrelin GHS1a receptor instead of one of the adenosine receptors.

Panel A. Saturation BRET curve in HEK-293T co-transfected with 1.5 μ g of the two cDNA corresponding to A₁R-cRLuc8 and A_{2A}R-nRLuc8 and with increasing amounts of cDNAs corresponding to A₁R-nVenus and A_{2A}R-cVenus (equal amounts of the two cDNAs). BRET_{max} was 35 \pm 2 mBU and BRET₅₀ was 16 \pm 3. BRET in cells expressing cRluc8 instead of A₁R-cRluc8 gave a linear-non saturable- signal.

Panel B. Comparison of BRET responses using complementary and non-complementary pairs, or replacing one adenosine receptor by the ghrelin GHS1a (gn) receptor. Data are mean \pm SD of three different experiments grouped as a function of the amount of BRET acceptor.

*** $p < 0.001$ respect to BRET in cells expressing adenosine receptors and hemi-Rluc8 and Venus proteins.

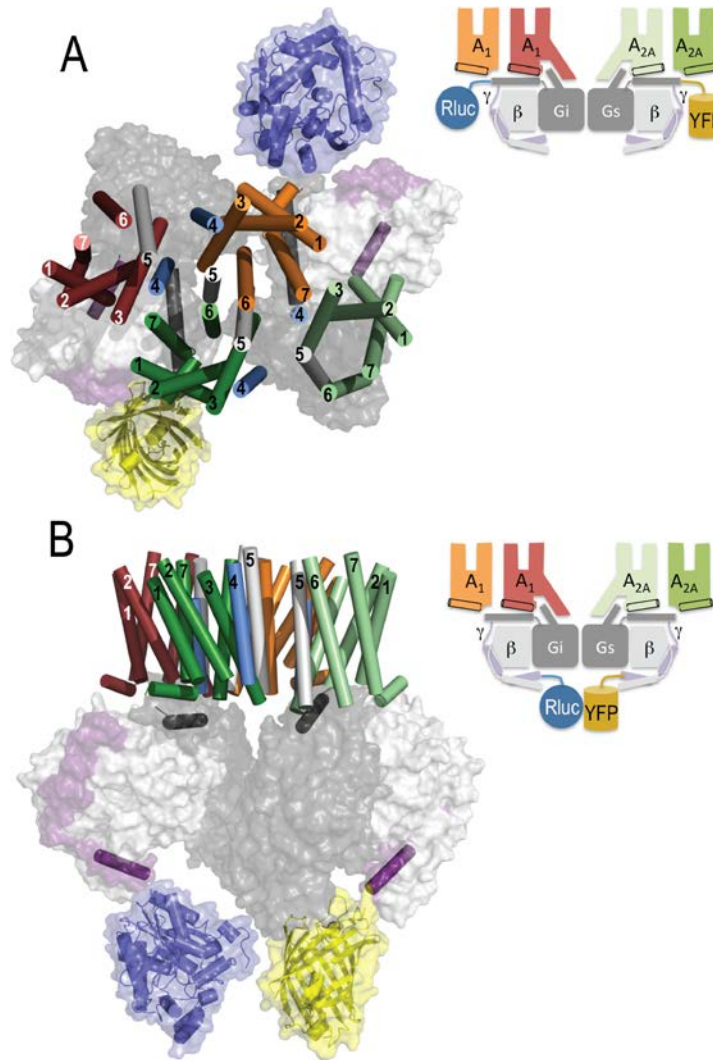


Figure S 6. Details of the relative position of Rluc and YFP in a receptor heterotetramer interacting with two G proteins

Computational-based model of G_s and G_i bound to the adenosine A_1R - $A_{2A}R$ heterotetramer. Rluc and YFP fused to the N-terminal domain of the G_α -subunits point toward different positions in space (A), whereas Rluc and YFP fused to G_γ -subunits are close (B). The color code of the proteins is depicted in the adjacent

schematic representations (TM4 and TM5 of GPCR protomers are in light blue and gray, respectively).

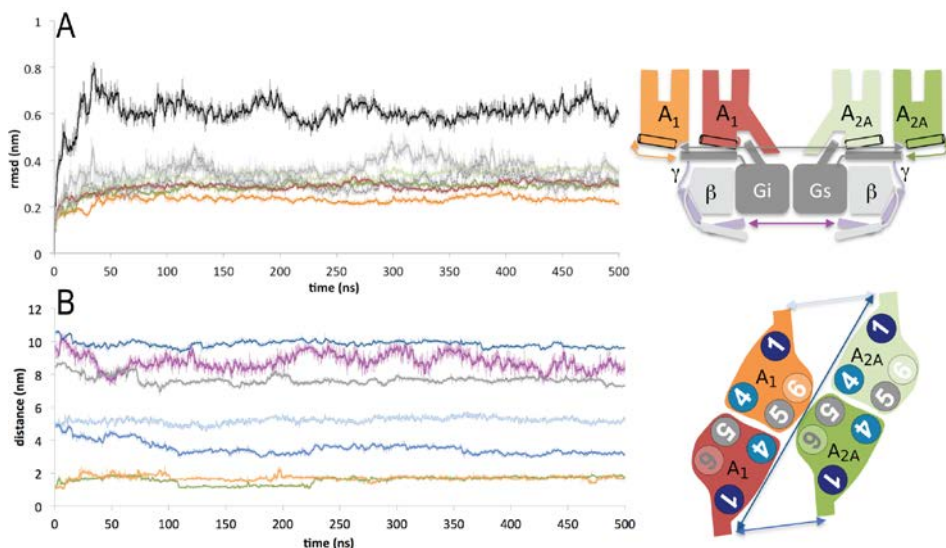


Figure S7. Molecular dynamics simulation of the adenosine A₁-A_{2A} receptor heterotetramer in complex with G_i and G_s

(A) Root-mean-square deviations (rmsd) on protein α -carbons of the whole system (black solid line), of the two A₁Rs (orange and red solid lines), of the two A_{2A}Rs (light and dark green solid lines), of G_i (gray solid line), and G_s (gray dotted line) throughout the MD simulation. This color scheme matches with the color of the different proteins depicted in the two adjacent schematic representations.

(B) Intermolecular distances between the N-terminal helices of the γ -subunit of G_i and G_s (magenta line), the N-terminal helices of the α -subunit of G_i and G_s (gray line), the N-terminal helix of the α -subunit of G_i and the C-terminal helix (Hx8) of inactive A₁R (orange line), the N-terminal helix of the α -subunit of G_s and the C-terminal Hx8 of inactive A_{2A}R (green line), the C-terminal Hx8 of A₁R and A_{2A}R (blue lines). These computed intermolecular distances are depicted as double arrows in the two adjacent schematic representations.

(A) Representative snapshots (20 structures collected every 25 ns) of the transmembrane domains of A_1R bound to G_i (red), G_i -unbound A_1R (orange), $A_{2A}R$ bound to G_s (dark green), and G_s -unbound $A_{2A}R$ (light green). TM helices 4 and 5 are highlighted in light blue and gray, respectively. Initial (at 0 ns, transparent cylinders) and final (at 500 ns, solid cylinders) snapshots of TM interfaces for homo- (TM4/5, within rectangles) and hetero- (TM5/6, within a circle) dimerization bundles. TM helices 4 (light blue), 5 (gray) and 6 (orange and green) are highlighted.

(B) Root-mean-square deviations (rmsd) on protein α -carbons of the four-helix bundles forming the TM5/6 interface (orange solid line), TM4/5 interface of A_1R (blue dotted line) and TM4/5 interface of $A_{2A}R$ (blue solid line) throughout the MD simulation.

(C) Contact maps of the TM4/5 interface (rectangles in panel A) in the A_1R or $A_{2A}R$ homodimer (left and right panels) and of the TM5/6 interface (circle in panel A) in the A_1R - $A_{2A}R$ heterodimer (middle panel). Darker dots show more frequent contacts.

(D) Detailed view of the extensive network of hydrophobic interactions (mainly of aromatic side chains) within the TM4/5 (left and right panels) and TM5/6 (middle panel) interfaces. The amino acids are numbered following the generalized numbering scheme of Ballesteros & Weinstein.^{37,38} This allows easy comparison among residues in the 7TM segments of different receptors.

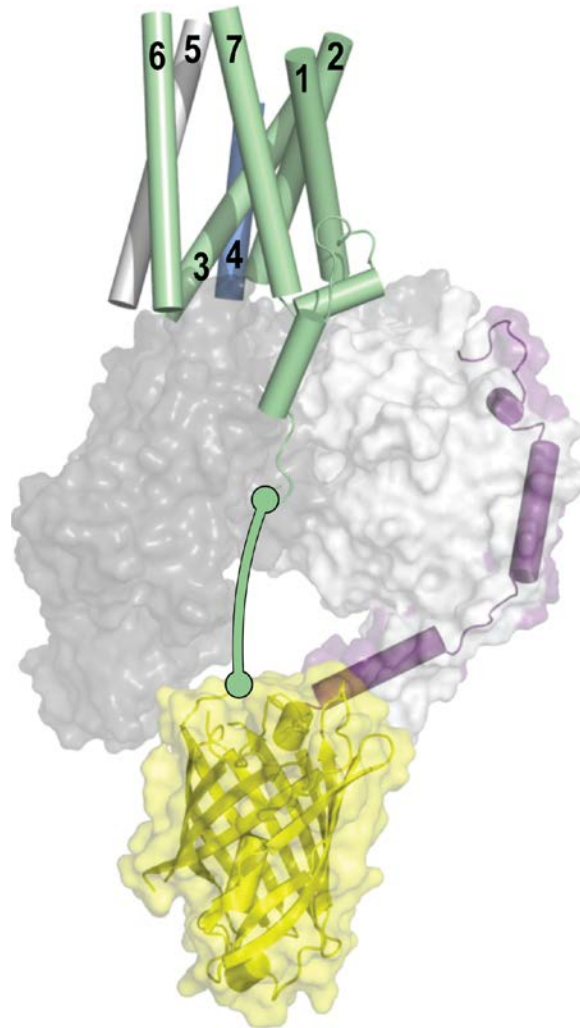


Figure S9. Positioning YFP in the C-tail of A_{2A}R

The complex between the A_{2A}R protomer (in light green) and G_s (α-subunit in dark grey and yellow, β-subunit in light grey, and γ-subunit in purple) was constructed from the crystal structure of β₂- in complex with G_s.³³ Although the exact conformation of the A_{2A}R C-tail (102 amino acids, Gln311-Ser412) cannot unambiguously be determined, its orientation was modeled as in the C-tail of squid rhodopsin³⁹ that contains the conserved amphipathic helix 8 that runs parallel to the membrane and an additional cytoplasmic helix 9. Thus, the C-tail of A_{2A}R expands (see solid light green line) points intracellularly toward the N-

termini of the γ -subunit as suggested for OXER.³² The laboratory of Kostenis has shown that the C-term of OXER, labeled with Rluc (OXER-Rluc), gets close to the N-term of the γ -subunit, labeled with GFP (γ -GFP). Analogously, we propose that YFP attached to the C-tail of $A_{2A}R$ is positioned near the N-termini of the γ -subunit (in purple).



Movie M1. Assembly of adenosine A_1 and A_{2A} receptors in complex with two G proteins and molecular dynamics simulations of the system

The assembly of G_s and G_i bound to the adenosine A_1R - $A_{2A}R$ heterotetramer was subjected to 500 ns of molecular dynamics simulations in a rectangular box containing the system, the lipid bilayer, explicit solvent and ions. A_1R protomers are in orange and red, $A_{2A}R$ protomers in light and dark green, G_α in white, G_β in gray, and G_γ in purple. For easier visualization of protomer-protomer interfaces, TMs 4 and 5 are highlighted in blue and white, respectively

3.8.6. References

- 1 Fung, J. J. *et al.* Ligand-regulated oligomerization of $\beta(2)$ -adrenoceptors in a model lipid bilayer. *The EMBO Journal* **28**, 3315-3328, (2009).
- 2 Albizu, L. *et al.* Time-resolved FRET between GPCR ligands reveals oligomers in native tissues. *Nature chemical biology* **6**, 587-594, (2010).
- 3 Manglik, A. *et al.* Crystal structure of the μ -opioid receptor bound to a morphinan antagonist. *Nature* **485**, 321-326, (2012).
- 4 Huang, J., Chen, S., Zhang, J. J. & Huang, X.-Y. Crystal Structure of Oligomeric $\beta(1)$ -Adrenergic G Protein-Coupled Receptors in Ligand-Free Basal State. *Nature structural & molecular biology* **20**, 419-425, (2013).
- 5 Lane, J. R. *et al.* A new mechanism of allostery in a G protein-coupled receptor dimer. *Nat Chem Biol* **10**, 745-752, (2014).
- 6 Viñals, X. *et al.* Cognitive Impairment Induced by Delta9-tetrahydrocannabinol Occurs through Heteromers between Cannabinoid CB₁ and Serotonin 5-HT_{2A} Receptors. *PLoS Biol* **13**, e1002194, (2015).
- 7 Kasai, R. S. & Kusumi, A. Single-molecule imaging revealed dynamic GPCR dimerization. *Current Opinion in Cell Biology* **27**, 78-86, (2014).
- 8 Hern, J. A. *et al.* Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules. *Proceedings of the National Academy of Sciences* **107**, 2693-2698 (2010).
- 9 Calebiro, D. *et al.* Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes with distinct dynamics and organization. *Proceedings of the National Academy of Sciences* **110**, 743-748 (2013).
- 10 Kasai, R. S. *et al.* Full characterization of GPCR monomer-dimer dynamic equilibrium by single molecule imaging. *The Journal of Cell Biology* **192**, 463-480, (2011).
- 11 Vilardaga, J.-P. *et al.* Conformational cross-talk between $\alpha 2A$ -adrenergic and μ -opioid receptors controls cell signaling. *Nat Chem Biol* **4**, 126-131, (2008).

- 12 Fribourg, M . *et al.* Decoding t he S ignaling o f a G PCR Heteromeric C omplex Reveals a U nifying M echanism o f Action of A ntipsychotic D rugs. *Cell* **147**, 1011 -1023, (2011).
- 13 González, S . *et al.* Circadian-Related H eteromerization o f Adrenergic a nd D opamine D₄ Receptors M odulates Melatonin Synthesis and Release in the Pineal Gland. *PLoS Biol* **10**, e1001347, (2012).
- 14 Navarro, G . *et al.* Interactions b etween Intracellular Domains a s Key D eterminants of t he Quaternary S tructure and F unction of R eceptor H eteromers. *The J ournal of Biological Chemistry* **285**, 27346-27359, (2010).
- 15 Ferré, S . *et al.* Building a ne w c onceptual framework f or receptor h eteromers. *Nature c hemical bi ology* **5**, 131 -134, (2009).
- 16 Venkatakrisnan A J, D . X ., Lebon G , T ate C G, S chertler GF, B abu M M. M olecular signatures of G -protein-coupled receptors. *Nature* **494**, 185-194 (2013).
- 17 Ciruela, F. *et al.* Presynaptic control of striatal glutamatergic neurotransmission b y a adenosine A₁-A_{2A} r eceptor heteromers. *The Journal of Neuroscience* **26**, 2080 (2006).
- 18 Orru, M . *et al.* Striatal P re- and P ostsynaptic P rofile of Adenosine A_{2A} R eceptor A ntagonists. *PLoS O NE* **6**, e16088, (2011).
- 19 Harms, G. S . *et al.* Single-Molecule I maging of L -Type Ca²⁺ Channels in Live Cells. *Biophysical Journal* **81**, 2639-2646, (2001).
- 20 Whorton, M . R . *et al.* A m onomeric G pr otein-coupled receptor i solated i n a h igh-density l ipoprotein pa rticle efficiently activates its G p rotein. *Proceedings of t he National A cademy of Sciences of t he United St ates of America* **104**, 7682-7687, (2007).
- 21 Ferré, S. *et al.* G Protein–Coupled Receptor Oligomerization Revisited: F unctional a nd P harmacological P erspectives. *Pharmacological Reviews* **66**, 413-434, (2014).
- 22 Fredholm, B. B., Ijzerman, A. P., Jacobson, K. A., Linden, J. & Müller, C. E. International U nion of Basic and Clinical Pharmacology. LXXXI. Nomenclature and Classification of Adenosine R eceptors—An U pdate. *Pharmacological Reviews* **63**, 1-34, (2011).

- 23 Carriba, P. *et al.* Detection of heteromerization of more than
two proteins by sequential BRET-FRET. *Nat Meth* **5**, 727-
733, (2008).
- 24 Gilchrist, A., Li, A. & Hamm, H. E. G α COOH-Terminal
Minigene Vectors Dissect Heterotrimeric G Protein
Signaling. *Science STKE* **2002**, pl1 (2002).
- 25 Cordoní, A., Navarro, G., Aymerych, M. S. & Franco, R.
Structures for G-Protein-Coupled Receptor Tetramers in
Complex with G Proteins. *Trends in Biochemical Sciences*
40, 548-551, (2015).
- 26 Mondal, S. *et al.* Membrane Driven Spatial Organization of
GPCRs. *Scientific Reports* **3**, 2909, (2013).
- 27 Okada, M. *et al.* Adenosine Receptor Subtypes Modulate
Two Major Functional Pathways for Hippocampal Serotonin
Release. *The Journal of Neuroscience* **21**, 628 (2001).
- 28 Cristóvão-Ferreira, S. *et al.* A(1)R–A(2A)R heteromers
coupled to G(s) and G(i/o) proteins modulate GABA
transport into astrocytes. *Purinergic Signalling* **9**, 433-449,
(2013).
- 29 Canals, M. *et al.* Adenosine A2A-Dopamine D2 Receptor-
Receptor Heteromerization: qualitative and quantitative
assessment by fluorescence and bioluminescence energy
transfer. *Journal of Biological Chemistry* **278**, 46741 -
46749 (2003).
- 30 Liu, W. *et al.* Structural Basis for Allosteric Regulation of
GPCRs by Sodium Ions(). *Science (New York, N.y.)* **337**,
232-236, (2012).
- 31 Martí-Renom, M. A. *et al.* Comparative Protein Structure
Modeling of Genes and Genomes. *Annual Review of
Biophysics and Biomolecular Structure* **29**, 291-325, (2000).
- 32 Blättermann, S. *et al.* A biased ligand for OXE-R uncouples
G α and G $\beta\gamma$ signaling within a heterotrimer. *Nat Chem Biol*
8, 631-638, (2012).
- 33 Rasmussen, S. G. F. *et al.* Crystal structure of the [bgr]2
adrenergic receptor-Gs protein complex. *Nature* **477**, 549-
555, (2011).
- 34 Chung, K. Y. *et al.* $\beta(2)$ adrenergic receptor-induced
conformational changes in the heterotrimeric G protein Gs.
Nature **477**, 611-615, (2011).

- 35 Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854, (2013).
- 36 Cordoní, A., Caltabiano, G. & Pardo, L. Membrane Protein Simulations Using AMBER Force Field and Berger Lipid Parameters. *Journal of Chemical Theory and Computation* **8**, 948-958, (2012).
- 37 Ballesteros, J. A. & Weinstein, H. In *Methods in Neurosciences* Vol. Volume 25 (ed C. Sealfon Stuart) 366-428 (Academic Press, 1995).
- 38 Isberg, V. *et al.* Generic GPCR Residue Numbers - Aligning Topology Maps Mapping The Gaps. *Trends in pharmacological sciences* **36**, 22-31, (2015).
- 39 Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453**, 363-367, (2008).

3.9. The transmission switch mechanism of allosteric modulation of the metabotropic glutamate 2 receptor

3.9.1. Background

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system of vertebrates and contributes excitatory input to as many as 80 to 90 % of central synapses.¹ It plays a major role in numerous physiological functions, such as learning and memory but also sensory perception, development of synaptic plasticity, motor control, respiration, and regulation of cardiovascular function. An imbalance in glutamatergic neurotransmission is believed to be at the center of various neurological and psychiatric diseases.^{2,3} Glutamate acts via activation of ionotropic or metabotropic glutamate receptors (iGlu or mGlu). The former includes ion channels such as NMDA, AMPA and kainate receptors responsible for fast excitatory transmission; whereas the latter are a family of eight class C GPCRs contributing to the fine-tuning of synaptic efficacy.^{2,4} Within this family, the mGlu₂ receptor is of particular importance for neuropharmacology as it is expressed presynaptically and

negatively modulates glutamate and GABA release. Glutamate hyperfunction is linked with disorders such as anxiety and schizophrenia⁵⁻⁷ hence, mGlu₂ receptor activation, is seen as a drug discovery approach to dampen these effects. Alternatively, blocking of mGlu₂ receptors is expected to have a beneficial effect on glutamate receptor hypofunction which may provide treatment for depression disorders⁸ or even be linked with cognition enhancement and possible therapy for Alzheimer's disease.⁹

The first generation of pharmacological tools for mGlu₂ receptors were conformationally constrained analogs of glutamate active as agonists at both mGlu₂ and mGlu₃ receptors.¹⁰ Whilst these molecules have progressed through preclinical¹¹ and clinical testing,^{12,13} various difficulties such as selectivity, brain penetration, and intellectual property space hamper drug discovery in this area. In recent years, there have been huge advances in the discovery of allosteric modulators that bind at less conserved allosteric sites and act in conjunction with the endogenous ligand.¹⁴ Allosteric modulation can decrease (negative allosteric modulator, NAM) or increase (positive allosteric modulation, PAM) the action (affinity and/or efficacy) of the orthosteric ligand.¹⁵ As such, PAMs have emerged as arguably the preferred approach to activate the receptor

with many chemical series being characterized in animal models providing preclinical proof of the PAM approach.^{16,17,18-26} To date, two mGlu₂ PAM molecules have advanced into clinical trials, AZD8529²⁷ and JNJ-40411813^{28,29} (also known as A-DX71149). Various key reference molecules are known in the field, including BINA (**1**),³⁰ JNJ-40068782 (**2**)³¹ and JNJ-46281222 (**3**).³² Meanwhile, inhibition of mGlu₂ receptors also started with development of orthosteric ligands such as the well-studied LY351495 from Eli Lilly,³³ and subsequently shifted to allosteric antagonists. F. Hoffmann-La Roche has been major players in the development of NAMs, with molecules such as Ro-676221 (**4**),⁴ Ro-4491533 (**5**),³⁴ and Ro-4995819 (**6**) being representative examples of their work. A number of these NAMs have been characterized *in vivo*.³⁴⁻³⁶ Molecule **6** (Ro-4995819/Decoglurant) has advanced into clinical trials,³⁷ and its use with other mGlu_{2/3} antagonists is reported for the treatment of autistic disorders.³⁸

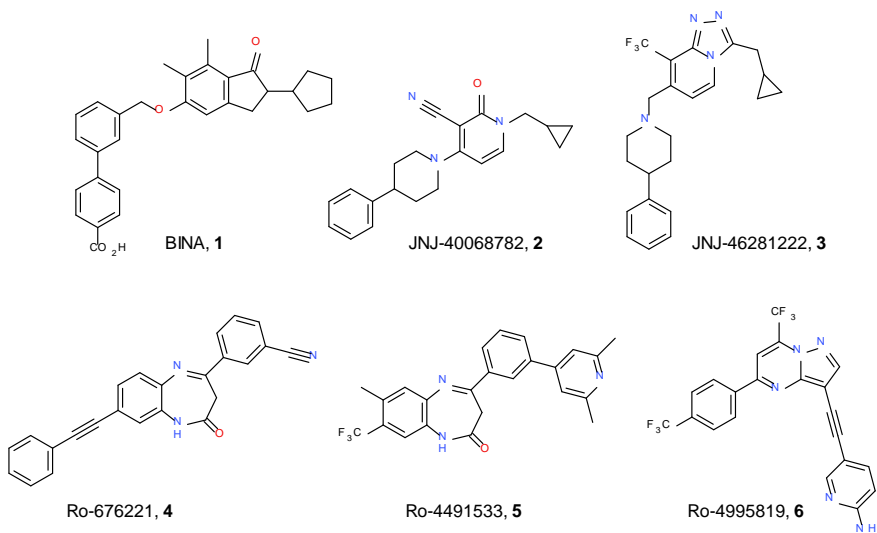


Figure 1. Selected mGlu₂ receptor PAMs (1 to 3) and NAMs (4 to 6) studied in this work.

The mGlu receptors exist as homodimers, which are covalently linked via a disulfide bond in the orthosteric extracellular binding domain.³⁹ Glutamate binding induces a conformational change in the receptor resulting in activation of the G protein and intracellular signaling.⁴⁰ Activation of the mGlu₂ receptor shifts the 7-TM dimer interface from TMs 4 and 5 in the inactive state to TM6 in the active state.⁴¹ Mutagenesis studies demonstrated that mGlu₂ allosteric modulators bind in an evolutionary conserved site in the upper half of the 7-TM similar to that of orthosteric ligands in Class A.⁴²⁻⁴⁶ Indeed, monomeric mGlu₂ receptors, either as an isolated 7-TM domain or in full-length, couple to G proteins upon activation

by a PAM alone.⁴⁷ Also, only one of the allosteric binding sites in the mGlu homodimer needs to be occupied by a PAM in order to achieve maximal potentiation, hence only one monomer is activated.⁴⁸ The recent Class C crystal structures of mGlu₁ and mGlu₅ receptors solved with NAMs have confirmed the 7-TM binding site.^{49,50} The crystallized ligands make few polar interactions, instead forming multiple hydrophobic and lipophilic contacts and providing steric complementarity with the receptor. However, there are some differences for these closely related mGlu receptors; Mavoglurant binds much deeper into the mGlu₅ receptor than FITM at mGlu₁.

Given the attraction of allosteric GPCR modulation in drug discovery research, here we perform a detailed study based upon mGlu₂ receptor experimental functional activity, binding data, site directed mutagenesis and computational structure activity relationships, docking and molecular dynamics (MD) simulations. The combined analysis defines how mGlu₂ allosteric modulators bind, and in particular, how the ligand exerts its allosteric functional effects. Up to now, these aspects have been largely unknown but the insight provided will provide new directions for drug discovery in this area. In addition, it will further our understanding of GPCR

allosteric modulation and the synergies between families and classes.

3.9.2. Results

Molecules 1 to 6 are functional allosteric modulators with an overlapping binding site

Molecules **1** to **6** were tested in mGlu₂ receptor functional assays aimed to detect positive or negative allosteric modulation (Table 1). Molecules **1**, **2** and **3** showed the characteristic profile of a PAM with pEC₅₀'s of 7.03, 6.90 and 8.09 respectively. These molecules showed no effect up to the concentration limit of 10 or 30 μM in the NAM assay. Correspondingly, molecules **4**, **5** and **6** showed mGlu₂ receptor NAM activity with pIC₅₀'s of 8.29, 8.57 and 8.60 respectively. In turn, these molecules showed no effect in the PAM assay up to the concentration limits. All PAMs and NAMs competed with the tritiated PAM [³H]JNJ-46281222 in the binding displacement assay, displaying pK_i's ranging from 7.22 to 8.33 and with Hill slopes approximately 1 suggesting a single population of binding sites. Hence, these molecules either augment or inhibit the

glutamate response at mGlu₂ receptors, by binding to an allosteric site, which is likely shared between PAMs and NAMs.

Table 1 mGlu₂ receptor activity and affinity of mGlu₂ positive and negative allosteric modulators determined by [³⁵S]GTPγS and [³H]JNJ-46281222 binding experiments using stably expressing hmGlu2-CHO cells

Compound		PAM	NAM	Binding
		pEC ₅₀ ^a	pIC ₅₀ ^b	pK _i ^c
1 , BINA	PAM	7.03 ± 0.14	< 5.0	7.22 ± 0.26
2 , JNJ-40068782	PAM	6.90 ± 0.10	< 5.0	7.58 ± 0.16
3 , JNJ-46281222	PAM	8.09 ± 0.23	< 4.52	8.33 ± 0.34
4 , Ro-676221	NAM	< 4.3	8.29 ± 0.21	7.96 ± 0.13
5 , Ro-4491533	NAM	< 4.3	8.57 ± 0.22	8.09 ± 0.16
6 , Ro-4995819	NAM	< 4.3	8.60 ± 0.07	7.56 ± 0.07

^a Functional activity of mGlu₂ receptor PAMs determined by the enhancement of glutamate (EC₂₀) induced [³⁵S]GTPγS binding. ^b Functional activity of mGlu₂ NAMs determined by the reduction of glutamate (EC₈₀) induced [³⁵S]GTPγS binding. ^c Affinity for the allosteric binding pocket of the mGlu₂ receptor as determined by [³H]JNJ-46281222 binding experiments. Data are shown as mean ± SD of at least three individual experiments performed in duplicate.

The binding site of NAMs and PAMs

To confirm the location of the binding site of both mGlu₂ NAMs and PAMs, we performed site directed mutagenesis experiments. All mGlu₂ receptor mutants confirmed proper orthosteric receptor binding and showed similar affinity for glutamate, as determined using [³H]LY341495 (data not shown). Results for PAMs **1**, **2** and **3** are presented in Table 2. Although several of the mutated amino acids are not important for activity, the mutagenesis experiments reveal that R635^{3.28a.32c}A (numbering as recommended for class A and class C GPCRs⁵¹), L639^{3.32a.36c}A, F643^{3.36a.40c}A, N735^{5.47a.47c}D, and W773^{6.48a.50c}A mutations affected the activity of PAMs. The H723^{5.34a.35c}V mutation affects the function of molecule **2** and F776^{6.51a.53c}A affects the function of molecules **1** and **2**. Mutagenesis experiments of NAMs are reported in Table 3. Clearly, F643^{3.36a.40c}A, L732^{5.43a.44c}A, W773^{6.48a.50c}A, F776^{6.51a.53c}A, and F780^{6.55a.57c}A mutations reduced the activity of NAMs **4**, **5** and **6**. The L639^{3.32a.36c}A mutation only significantly reduced the activity of NAMs **4** and **6**, whereas I693^{4.56a.46c}M, D725^{5.36a.37c}A and V798^{7.42a.36c}A only affected the activity of **6**. Figure 2 shows concentration-response curves for the mutants that produce the largest effect on the activity of each of the three NAMs (see

supplementary information Figure S1 for additional curves). Figure 3 also shows that NAM 5 affects the glutamate concentration-response curve similar for both WT and mutant F776^{6.51a.53c}A mGlu₂ receptors (we previously described similar results for PAMs⁴²).

Only L639^{3.32a.36c}, F643^{3.36a.40c}, W773^{6.48a.50c} and F776^{6.51a.53c} are important for the action of PAMs and NAMs, suggesting that within the shared binding site, there are both common and specific interactions for the type of allosteric modulator. Interestingly, R635^{3.28a.32c} and N735^{5.47a.47c} affected only PAMs but not NAMs, whereas F780^{6.55a.57c}A only affected NAMs and not PAMs.

Table 2. Effect of mGlu₂ receptor mutations on activity of PAMs as determined by [³⁵S]GTPγS binding assay in the presence of an EC₂₀ glutamate concentration (4 μM).

Mutant	PAM compound		
	1, BINA	2, JNJ-40068782	3, JNJ-46281222
	pEC ₅₀	pEC ₅₀	pEC ₅₀
Transient WT	7.11 ± 0.30	7.08 ± 0.13	8.22 ± 0.23
Stable WT	7.03 ± 0.14	6.88 ± 0.13	8.09 ± 0.23
R635 ^{3.28a.32c} A	6.21 ^{a***}	6.42 ± 0.03 ^{***}	n.c.
R636 ^{3.29a.33c} A	n.c.	n.c.	n.c.
L639 ^{3.32a.36c} A	5.89 ± 0.01 ^{a***}	6.27 ± 0.16 ^{***}	n.c.
F643 ^{3.36a.40c} A	5.85 ± 0.08 ^{a***}	5.64 ± 0.18 ^{***}	6.32 ± 0.25 ^{a***}
S644 ^{3.37a.41c} A	6.99 ± 0.15	6.7 ± 0.22	n.c.
S688 ^{4.51a.41c} L	6.93 ± 0.33	6.90 ± 0.44	7.69
G689 ^{4.52a.42c} V	6.20 ± 0.37 [*]	6.27 ± 0.20 ^{***}	7.21 ± 0.30 [*]
I693 ^{4.56a.46c} M	n.c.	n.c.	n.c.
V700 ^{4.63a.53c} L	n.c.	n.c.	n.c.
H723 ^{5.34a.35c} V	n.c.	6.26 ± 0.06 ^{***}	n.c.
D725 ^{5.36a.37c} A	n.c.	n.c.	n.c.
M728 ^{5.39a.40c} A	n.c.	n.c.	n.c.
S731 ^{5.42a.43c} A	n.c.	n.c.	n.c.
L732 ^{5.43a.44c} A	7.80 ± 0.03	7.03 ± 0.06	7.31 ± 0.33
N735 ^{5.47a.47c} D	5.13 ^{a***}	5.52 ^{a***}	6.67 ± 0.23 ^{a***}
V736 ^{5.48a.48c} A	n.c.	n.c.	n.c.
W773 ^{6.48a.50c} A	< 5.0 ^{***}	< 5.0 ^{***}	7.11 ^{a,b***}
F776 ^{6.51a.53c} A	6.70 ± 0.06	6.38 ± 0.23 ^{***}	n.c.
F780 ^{6.55a.57c} A	n.c.	n.c.	n.c.
V798 ^{7.42a.36c} A	n.t.	n.t.	n.t.

* p < 0.05, ** p < 0.01, *** p < 0.001 significantly different from value obtained for transiently transfected WT mGlu₂ receptor. Determined using one-way ANOVA with Dunnett's post-test. Data represent the mean ± SD of at least three individual experiments performed in triplicate. ^aFor one or two experiments, pEC₅₀ was < 5; ^bn=2. n.c.: No change in response upon testing compound at 2 concentrations (concentrations chosen are equivalent to concentration producing half-maximal or maximal response as determined for these compounds on the WT receptor); n.t.: Not tested

Table 3. Effect of mGlu₂ mutations on activity of NAMs as determined by [³⁵S]GTPγS binding assay in the presence of an EC₈₀ glutamate concentration (60 μM).

Mutant	NAM compound		
	4, Ro-676221	5, Ro-4491533	6, Ro-4995819
	pIC ₅₀	pIC ₅₀	pIC ₅₀
Transient WT	8.18 ± 0.28	8.69 ± 0.18	9.00 ± 0.21
Stable WT	8.29 ± 0.21	8.57 ± 0.22	8.60 ± 0.07
R635 ^{3.28a.32c} A	8.08 ± 0.09 ^a	8.91 ± 0.50 ^a	n.t.
R636 ^{3.29a.33c} A	8.29 ± 0.24	8.75 ± 0.28	8.68 ± 0.11
L639 ^{3.32a.36c} A	7.79 ± 0.28 ^{**}	8.40 ± 0.25	8.41 ± 0.09 ^{***}
F643 ^{3.36a.40c} A	7.39 ± 0.21 ^{***}	7.50 ± 0.20 ^{***}	7.17 ± 0.07 ^{***}
S644 ^{3.37a.41c} A	8.32 ± 0.35	8.77 ± 0.37	n.t.
S688 ^{4.51a.41c} L	8.17 ± 0.12	8.43 ± 0.35	n.t.
I693 ^{4.56a.46c} M	8.06 ± 0.09	8.67 ± 0.03	8.47 ± 0.11 ^{***}
V700 ^{4.63a.53c} L	8.26 ± 0.10 ^a	8.65 ± 0.23	n.t.
H723 ^{5.34a.35c} V	8.30 ± 0.29	8.61 ± 0.30	8.62 ± 0.08
D725 ^{5.36a.37c} A	8.15 ± 0.16	8.38 ± 0.10	8.07 ± 0.14 ^{***}
M728 ^{5.39a.40c} A	8.55 ± 0.07	8.49 ± 0.07	9.23 ± 0.04
S731 ^{5.42a.43c} A	8.55 ± 0.16	8.71 ± 0.23	9.14 ± 0.08
L732 ^{5.43a.44c} A	7.64 ± 0.28 ^{***}	7.21 ± 0.15 ^{***}	8.33 ± 0.09 ^{***}
N735 ^{5.47a.47c} D	8.21 ± 0.54 ^a	8.64 ± 0.54 ^a	8.67 ± 0.05 ^a
V736 ^{5.48a.48c} A	8.00 ± 0.02 ^a	9.12 ± 0.23 ^a	n.t.
W773 ^{6.48a.50c} A	< 5	6.73 ± 0.09 ^{***}	6.75 ± 0.06 ^{***}
F776 ^{6.51a.53c} A	7.46 ± 0.64 ^{***}	7.35 ± 0.14 ^{***}	6.47 ± 0.09 ^{***}
F780 ^{6.55a.57c} A	7.58 ± 0.21 ^{***}	7.77 ± 0.12 ^{***}	7.54 ± 0.02 ^{***}
V798 ^{7.42a.36c} A	7.98 ± 0.20	8.45 ± 0.57	8.57 ± 0.26 ^{**}

* p < 0.05, ** p < 0.01, *** p < 0.001 significantly different from value obtained for transiently transfected WT mGlu₂ receptor. Determined using one-way ANOVA with Dunnett's post-test. Data represent the mean ± SD of at least three individual experiments performed in triplicate, except ^a which were two individual experiments were performed in triplicate. n.t. Not tested.

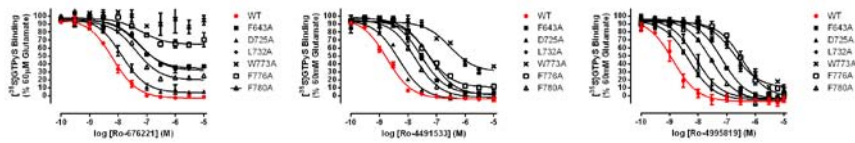


Figure 2. Concentration response curves of the most active mutants for each mGlu₂ NAM, **4** Ro-676221, **5** Ro-4491533, and **6** Ro-4995819. Further mGlu₂ receptor mutant NAM concentration response curves are shown in supporting information, Figure S1.

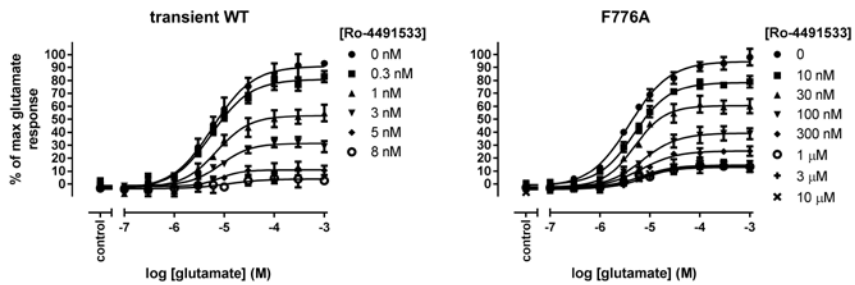


Figure 3. Cooperativity of NAM **5** Ro-4991533 on WT and mutant F776^{6.51a.53c}A mGlu₂ receptors.

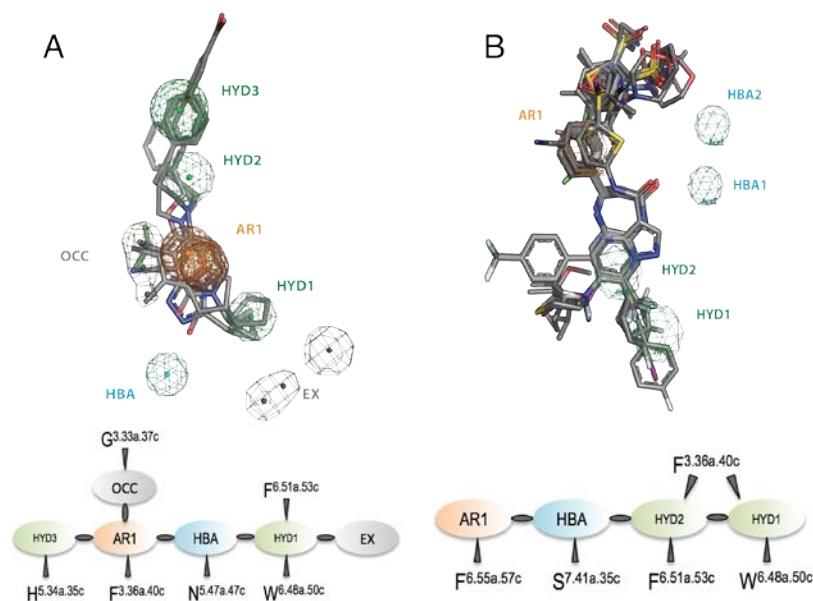


Figure 4. Pharmacophore models of PAMs (A) and NAMs (B) of the mGlu₂ receptor. The structural features of PAMs (panel A, top) are: an aromatic feature (AR1, in orange), a hydrogen bond acceptor group (HBA, cyan), three hydrophobic sites (HYD1-HYD3, green), an occupancy feature beneficial for activity (OCC, black), and an excluded feature detrimental for activity (EX, black). Schematic representation of the pharmacophore model of PAMs (panel A, bottom) and the predicted amino acids in the mGlu₂ receptor, determined by a combination of MD simulations (Figure 5) and site-directed mutagenesis experiments (Table 2). The structural features of NAMs (panel B, top) are: an aromatic feature (AR1, in orange), hydrogen bond acceptor groups (HBA1-HBA2, cyan), and two hydrophobic sites (HYD1-HYD2, green). Schematic representation of the pharmacophore model of NAMs (panel B, bottom) and the amino acids involved in the interaction with NAMs (Figure 5 and Table 3).

PAM and NAM Structure Activity Relationships

Using a systematic approach, we developed 3D pharmacophore models for PAMs and NAMs. Datasets of PAM and NAM molecules were compiled, molecules were converted to their 3D conformations and overlay models, including pharmacophoric features, were generated to separate known actives from inactive (see Methods). These pharmacophore models provide information about the Structure Activity Relationships (SAR) that are important for identifying active mGlu₂ receptor PAMs and NAMs, and hence what features of the molecules may interact with the receptor and what structural modifications can result in a loss of activity.

The PAM pharmacophore shows a clear structural overlap of the molecules with key shared features (see Figure 4A). *i.* A central aromatic feature (AR1) that orients the molecules via the π electrons through aromatic-aromatic or aromatic-hydrophobic interactions (see below). *ii.* A key hydrogen bond acceptor feature (HBA) (satisfied by the carbonyl group of the indanone or pyridone moieties in **1** and **2**, or by the sp² nitrogen in the triazolopyridine group of **3**). *iii.* A hydrophobic feature (HYD1) located at one end of the molecule, and two hydrophobic features (HYD2-HYD3) located at the other end that shows more variability of substituents.

iv. Active compounds commonly contain hydrophobic groups (methyl, cyano, trifluoromethyl, or chloro) on the AR scaffold. Thus, an additional occupancy feature (OCC) that is beneficial for activity was added to the pharmacophore model. *v.* In contrast, increasing the size of the HYD1 feature was detrimental for activity. Thus, an excluded feature (EX) was added to this part of the molecule. This pharmacophore is consistent with previous overlay hypotheses and has been used to understand SAR and develop new mGlu₂ PAM chemical series.^{23,52}

The NAM pharmacophore aligned molecules from the benzodiazapinones series (**4** and **5**) and also pyrazolopyrimidines (**6**) that have more structural diversity and hence a more divergent structural overlap than for PAMs. However, the scaffolds of the two different series were overlaid and they share common pharmacophoric features (Figure 4B). *i.* An aromatic feature (AR1). *ii.* A hydrogen bond acceptor feature (represented by HBA1 and HBA2) that is provided by the carbonyl group of the benzodiazapinone scaffold or by the amide carbonyl group (present in some of the pyrazolopyrimidine NAMs). Whilst this HBA feature is satisfied by all benzodiazapinones, the pyrazolopyrimidines do not always contain an amide carbonyl in this region (acetylenic

spacers can also be active). *iii*. Two hydrophobic features (HYD1-HYD2) located at one end of the molecule.

Binding mode of PAMs and NAMs at mGlu2 receptor

To better understand the affinity and selectivity of PAMs and NAMs for the mGlu₂ receptor, we docked PAMs **1-3** into the “active-like” model of mGlu₂ receptor in complex with Gi and NAMs **4-6** into the “inactive” model of mGlu₂ receptor (see Methods). Three independent unbiased 1 μs MD simulations were used to study the stability of the different docking solutions for each molecule. The simulations were also analyzed to elucidate details of the mechanism of action.

The binding mode hypotheses of PAMs **1-3** are robust and fluctuate little along the MD simulations (Figure 5 and Figure S2), with average ligand RMSD < 0.3 nm relative to the initial docking pose. They are also consistent with the SAR analysis. The carbonyl group (HBA in the pharmacophore model, Figure 4A) of PAM **1** acts as a hydrogen bond acceptor in the hydrogen bond with N735^{5.47a.47c}, the indanone ring (AR1) forms a parallel-displaced aromatic-aromatic interaction with F 643^{3.36a.40c}, the cyclopentyl moiety (HYD1) expands toward the intracellular side in interacting

with W773^{6.48a.50c} without reaching the EX feature, the methyl groups (OCC) enter into a small hydrophobic cavity between TMs 3 and 5 formed by G640^{3.33a.37c} (CAV1), the first aromatic ring of the biphenyl group (HYD2) expands toward the extracellular side interacting with L639^{3.32a.36c}, and the carboxylic acid forms an ionic interaction with R635^{3.38a.32c} (Figures 4A and 5A). Accordingly, the L639^{3.32a.36c}A, F643^{3.36a.40c}A, R635^{3.38a.32c}A, N735^{5.47a.47c}D, and W773^{6.48a.50c}A mutations have a significant effect on PAM **1** activity to mGlu₂ receptor (Table 2). The binding of PAMs **2** and **3** to mGlu₂ receptor follows similar trends as revealed by the MD simulations. The carbonyl group of **2** or the nitrogen atom of the triazole ring of **3** (HBA) form a hydrogen bond with N735^{5.47a.47c}, the pyridone ring of **2** or the triazolopyridine ring of **3** (AR1) interacts with F643^{3.36a.40c}, the common short cyclopropyl moiety (HYD1) interacts with W773^{6.48a.50c} and F776^{6.51a.53c}, and the cyano group of **2** or the trifluoromethyl group of **3** (OCC) enter the cavity CAV1 delimited by G640^{3.33a.37c} (Figures 5B and 5C). These binding modes are compatible with the mutagenesis experiments (Table 2). The common 4-phenylpiperidine (HYD2-3) substituents of PAMs **2** and **3** are directed towards the extracellular side. However, the extra flexibility of **3** compared to **2**, due to the

additional methylene group linking AR1 and HYD2, means PAM **2** can interact with L639^{3.32a.36c} and the H723^{5.34a.35c}-R635^{3.38a.32c} pair whereas PAM **3** does not contact these amino acids. Accordingly, the L639^{3.32a.36c}A, R635^{3.38a.32c}A, and H723^{5.34a.35c}V mutations have a significant effect on PAM **2** activity at mGlu₂ receptors and have no effect on **3** (Table 2). Based on these models, Figure 4 A (bottom) shows a schematic representation of the pharmacophore hypothesis and the predicted amino acids in the 7-TM domain of mGlu₂ receptor involved in the interaction with the ligands.

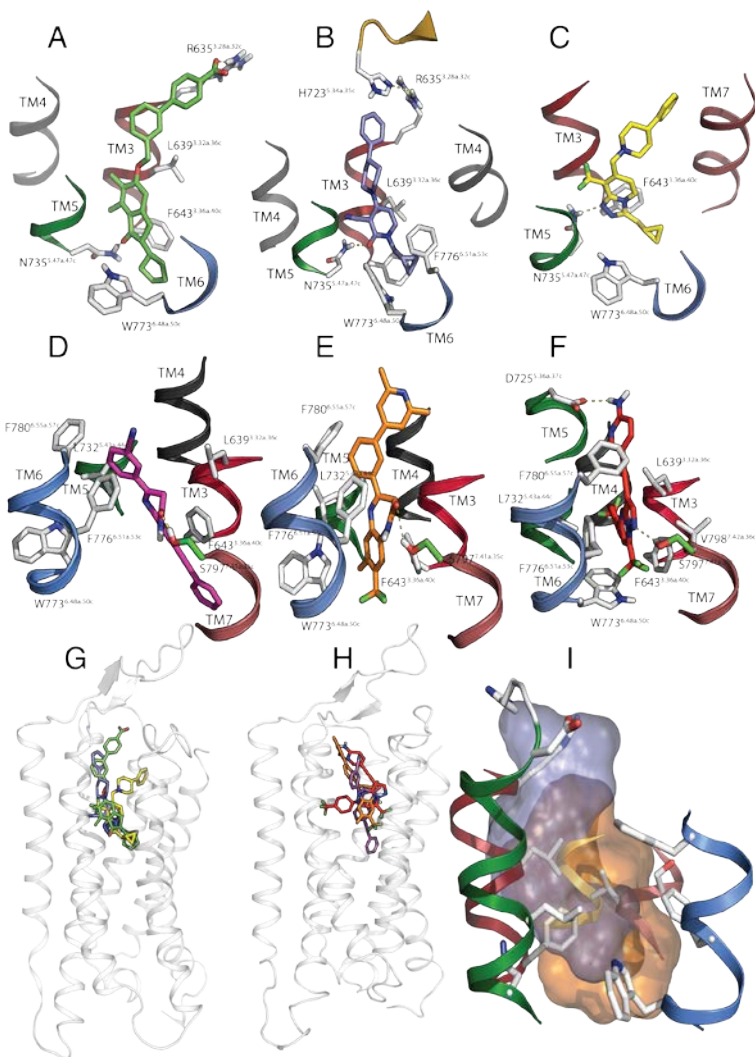


Figure 5. Detailed view of the binding mode of PAMs **1** (panel A, in green), **2** (panel B, blue), and **3** (panel C, yellow) and NAMs **4** (panel D, magenta), **5** (panel E, orange), and **6** (panel F, red) to the mGlu₂ receptor. Amino acids involved in the binding of ligands, as determined by site-directed mutagenesis experiments reported in Tables 2 and 3 are shown in white, whereas S797^{7.41a.35c} that was reported for the binding of NAMs to mGlu₅ receptor⁵³ is shown in green. Overlay of the binding modes of PAMs **1-3** (panel G) and NAMs **4-6** (panel H). Comparison of the binding cavities of PAMs (cyan) and NAMs (orange) at the mGlu₂ receptor (panel I).

Figure 5 (panels D-F) shows the binding mode of two different chemical series of NAMs: benzodiazapinones (NAMs **4** and **5**) and pyrazolopyrimidines (NAM **6**). All three NAMs bind close in front of TM6 with the central scaffold interacting with the aromatic W773^{6.48a.50c}, F776^{6.51a.53c} and F780^{6.55a.57c} amino acids. Accordingly, mutation of these amino acids to Ala has a significant effect on the activity of NAMs **4-6** at the mGlu₂ receptor (Table 3). For the interaction of NAM **5**, the carbonyl group (HBA1 in the pharmacophore model) of the benzodiazapinone moiety interacts with the side chain hydroxyl of S797^{7.41a.35c}, the aromatic moiety of the benzodiazapinone group (HYD2) forms a parallel-displaced aromatic-aromatic interaction with F643^{3.36a.40c} and a T-shaped aromatic-aromatic interaction with W773^{6.48a.50c}, the non-aromatic moiety of the benzodiazapinone group interacts with L732^{5.43a.44c} and F776^{6.51a.53c}, the trifluoromethyl group (HYD1) enters into a small hydrophobic cavity between TMs 6 and 7 (CAV2), the phenyl ring (AR1) forms an aromatic-aromatic interaction with F780^{6.55a.57c}, and the terminal pyridine group extends toward the extracellular side (Figure 5E). NAM **4** contains a phenylacetylene in place of the trifluoromethyl group of **5** that is too large to fit in the same CAV2 cavity between TMs 6 and 7 where the trifluoromethyl

group is placed. We observed that this phenylacetylene moiety in NAM **4** could enter deeper into the receptor in a similar manner as seen for the phenylacetylene moiety in the crystal structure of Mavoglurant bound to mGlu5 receptor.⁴⁹ This different location of HYD1 in NAMs **4** and **5** does not impede the other pharmacophoric elements to form comparable interactions, in agreement with mutagenesis experiments (Table 3). Thus, the common benzodiazapinone moiety (HBA and HYD2) of **4** interacts with S797^{7.41a.35c}, F643^{3.36a.40c}, L732^{5.43a.44c}, and W773^{6.48a.50c} and F776^{6.51a.53c}, the phenyl ring (AR1) with F780^{6.55a.57c}, and the cyano group expands toward the extracellular side (Figure 5D). Compound **6** contains two trifluoromethyl groups. Based on our findings, it is reasonable to assume that they would bind in the small hydrophobic cavities CAV1 between TMs 3 and 5 and CAV2 between TMs 6 and 7 in a similar manner as **3** and **5**, respectively (see above). Therefore, NAM **6** could bind in two different orientations locating each of the CF₃ groups in CAV1 or CAV2, respectively. Only the mode of interaction in which the phenyl-CF₃ group binds in CAV1 and the CF₃ group attached to the pyrazolopyrimidine ring binds in CAV2, depicted in Figure 5F, remained unchanged during the MD simulation and fulfills the results of the mutagenesis experiments

(see below). In this binding mode, the nitrogen atom (HBA) of the pyrazole ring interacts with S797^{7.41a.35c}, the pyrazolopyrimidine ring forms a parallel-displaced aromatic-aromatic interaction with F776^{6.51a.53c} and a T-shaped aromatic-aromatic interaction with W773^{6.48a.50c}, the phenyl-CF₃ moiety expands towards CAV1 and the phenyl ring interacts with F643^{3.36a.40c} in a parallel-displaced conformation, the triple bond and the aminopyridine group point towards the extracellular side interacting with F780^{6.55a.57c} (Figure 5F). We proposed that the 2-aminopyridine group of **6** also interacts with D725^{5.36a.37c} which would be lost upon Ala mutation which was confirmed experimentally (Table 3). The D725^{5.36a.37c}A only affects **6** because it is the only NAM with a binding mode which remains vertically aligned to TMs 5 and 6.

We tested all PAMs and NAMs **1-6** in mGlu receptor selectivity assays, see supporting information. Compounds were tested for activation or inhibition activity versus all 8 members of the mGlu family. PAMs **1** to **3** only showed activity at mGlu₂, and no activity versus other mGlu receptors. NAMs showed no activity in assays for activation of the mGlu receptor family, however, as well as showing inhibition of mGlu₂ activity, they also showed antagonistic activity at mGlu₃ receptors. This curious lack of selectivity of

mGlu2 NAMs versus mGlu3 receptors has been seen previously⁸ although without providing a structural explanation. The N735^{5.47a.47c}D mutation affects the activity of PAMs **1-3** but not NAMs **4-6** (Tables 2-3); hence NAMs do not interact with this amino acid. In contrast, N735^{5.47a.47c} forms a crucial hydrogen bond (HBA, Figure 4A) with PAMs. In mGlu3 receptors this amino acid is an aspartate (D) and cannot provide an H-bond donor in its ionized state therefore diminishing activity of mGlu₂ PAMs for the mutant and resulting in total selectivity versus WT mGlu₃. According to our binding mode hypotheses, this amino acid is key for the observed selectivity difference of PAMs and NAMs versus mGlu₃ and the mechanism of mGlu₂ receptor activation by PAMs.

3.9.3. Discussion

We have addressed the comparison of the binding of NAMs and PAMs and their mechanism of negative and positive allosteric modulation at the mGlu₂ receptor. The binding modes for PAMs and NAMs have been determined by a combination of experiment and computational methods. Mutants such as R 636^{3.29a.33c}A, F643^{3.36a.40c}A, H 723^{5.34a.35c}V, L 732^{5.43a.44c}A, W 773^{6.48a.50c}A, F780^{6.55a.57c}A, V 798^{7.42a.36c}A, I693^{4.56a.46c}M, D 725^{5.36a.37c}A and F776^{6.51a.53c}A were important for the activity of NAMs. In addition, we observed differences between residues important for PAMs and NAMs. Mutations R 636^{3.29a.33c}A, I 693^{4.56a.46c}M, F 780^{6.55a.57c}A and V798^{7.42a.36c}A had a large effect on the activity of NAMs but no effect on the PAMs. On the other hand, mutations L639^{3.32a.36c}A, S644^{3.37a.41c}A, S 688^{4.51a.41c}L/G689^{4.52a.42c}V and N 735^{5.47a.47c}D, did not affect the action of the NAMs but had a pronounced effect on the PAMs. Thus, our results show that the allosteric site is placed within the 7-TM domain, in a similar position as the NAMs of mGlu₁ and mGlu₅ observed in the crystal structures^{49,50,54} which also resembles the orthosteric site in class A GPCRs. Similarly to class A, in which a agonist and antagonist binding overlaps in the

orthosteric site but form different sets of interactions⁵⁵, the allosteric binding site of NAMs and PAMs at mGlu₂ receptors also overlaps but with significant differences (Figures 5G-5I). While the binding site of PAMs is close to TMs 3-5, NAMs bind in front of, and parallel to, TM6. Longer PAMs, such as **1**, become more flexible as they reach the extracellular region of the 7-TM, and move away from TMs 5 and 6 instead bending towards TM 3 and extracellular loop 2 (ECL2). In addition, NAMs expand deeper toward the intracellular side than PAMs.

GPCRs are dynamic proteins that permit rapid small-scale structural fluctuations and pass through an energy landscape to adopt a number of conformations.⁵⁶ MD simulations are being used more and more to study GPCR function due to the importance of generating ensembles of energetically accessible conformations^{57,58}. Thus, to study mechanisms of receptor inactivation by NAMs and receptor activation by PAMs, we used unbiased MD simulations in the microsecond time-scale to explore conformational changes at the receptor. Because families A and C of GPCRs bind the same type of G proteins, maintain the spatial conservation of the TM helices, and share the position of the binding site for orthosteric and allosteric ligands within the 7-TM domain, we assumed similar

mechanisms of receptor (in)activation, as previously tested for family B.^{59,60} Thus, we suggest that (in)activation of mGlu₂ receptors involves rearrangement of an analogous ‘transmission switch’ in TMs 3, 5, and 6 as described for class A,⁶¹ involving positions 3.40a, 5.50a and 6.44a, and for class B,⁶⁰ involving positions 3.40a.44b and 6.44a.49b. This ‘transmission switch’ was proposed for the highly studied β_2 -adrenergic receptor based on the fact that a hydrogen bond interaction between agonists and TM5, stabilizes a receptor conformation that includes an inward movement of TM5 at the highly conserved P^{5.50a}. The proline movement induces steric competition with a bulky hydrophobic side chain at position 3.40a, triggering a small counterclockwise rotation of TM3 when viewed from the extracellular side.⁶² Finally, the rotation of TM3 repositions the side chain of F^{6.44a}, facilitating the outward movement of TM6 for receptor activation and G protein binding.⁵⁵ Moreover, it is known that the initial agonist-induced structural changes of the receptor (‘trigger switch’) responsible for rearrangement of the ‘transmission switch’, are a hydrogen bond interaction between agonists and S^{5.46a} in β_1 -⁶³ and β_2 -adrenergic receptors,⁵⁵ the movement of W^{6.48a} in metarhodopsin II⁶⁴ and

A_{2A},⁶⁵ or a conformational toggle switch of the side chain of the amino acid at position 3.36a.⁶⁶ Thus, we analyzed the MD simulations of in active- and active-like mGlu2 receptor to study these 'switches' by monitoring the rotation and displacement of these key side chains in TMs 3, 5 and 6, and the influence of the NAMs or PAMs in the distribution of dihedral χ_1 angles in the MD trajectories (Figures 6-7).

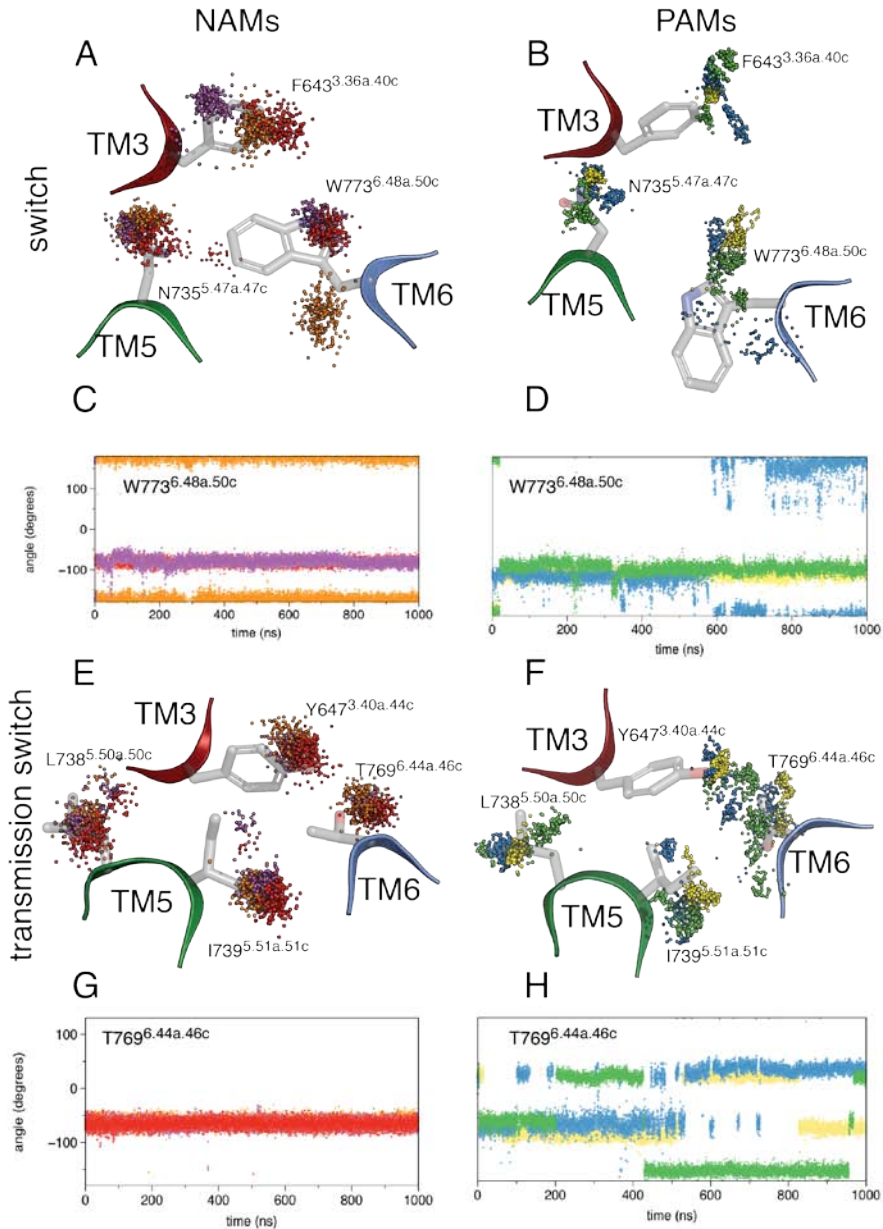


Figure 6 . Distribution of the positions of representative atoms (CZ atom of F643^{3.36a.40c}, the ND2 atom of N735^{5.47a.47c}, and the NE1 atom of W773^{6.48a.50c} (panels A-B), and the OH atom of Y647^{3.40a.44c}, CD2 atom of L738^{5.50a.50c}, CG2 atom of I739^{5.51a.51c}, and OG1 atom of T769^{6.44a.46c} (panels E-F)) of the amino acids of mGlu₂ receptor at the homologous positions of the ‘trigger switch’ and ‘transmission switch’ of class A, respectively, during MD simulations of the

“active-like” model of mGlu₂ receptor in complex with Gi and PAMs **1-3** (panels B, F) and of the “inactive” model of mGlu₂ receptor in complex with NAMs **4-6** (panels A, E). Evenly spaced snapshots extracted from the 1 μs of unbiased MD simulations are depicted. PAMs **1-3** are shown in green, blue and yellow and NAMs **4-6** are shown in magenta, orange and red, respectively. Time-evolution of the χ_1 rotamer of W773^{6.48a.50c} and T769^{6.44a.46c} for NAMs **4-6** (panels C, G) and PAMs **1-3** (panels D, H) during the MD simulations (line color indicates the simulation for each separate ligand bound to the receptor).

Figure 6 E illustrates the position of the side-chains of Y647^{3.40a.44c}, L738^{5.50a.50c}, I739^{5.51a.51c} and T769^{6.44a.46c} during the MD simulations for NAMs **4-6**. These amino acids are at the homologous positions of the ‘transmission switch’ of class A. Clearly, all NAMs stabilized similar conformations of these ‘transmission switch’ amino acids. Most obviously, T769^{6.44a.46c} moves so little throughout the simulations of all three NAMs that the separate dihedral angle distributions in Figure 6 G are indiscernible. This is in contrast to PAMs where it flips for each ligand throughout the simulations (Figure 6 H). Figure 6 A also shows similar analysis for the ‘trigger switch’ amino acids: F643^{3.36a.40c}, N735^{5.47a.47c}, and W773^{6.48a.50c}. Because NAMs do not interact with N735^{5.47a.47c}, no differences in TM5 are observed. In contrast, as shown by mutagenesis and docking experiments, all

NAMs form aromatic-aromatic interactions with F643^{3.36a.40c} (Table 3 and Figure 5). Thus, small differences are observed in the F643^{3.36a.40c} side chain (Figure 6A). A more significant difference is found at W773^{6.48a.50c}. Notably, NAM **5** that contains an extra methyl group relative to **6**, keeps the conformation of W773^{6.48a.50c} outside the 7-TM bundle, as observed in the crystal structures of the Mavoglurant-mGlu₅ receptor⁴⁹ and F ITM-mGlu₁ receptor⁵⁰ complexes. In contrast, NAMs **4** (due to the different position of its phenylacetylene moiety, Figure 5), and **6** (due to the absence of the methyl group) permit a conformational switch of W773^{6.48a.50c} towards the inside of the 7-TM bundle, as also observed in more recent mGlu₅ allosteric ligand crystal structures.⁵⁴ This is also reflected in the χ_1 rotamer conformation of W773^{6.48a.50c} (Figure 6C). NAMs **4** and **6** favor the *gauche+* conformation (inside the bundle) while NAM **6** favors the *trans* conformation (outside the bundle) of W773^{6.48a.50c}. Thus, the conformation of W773^{6.48a.50c} depends on the chemical structure of the NAM. Still, it is important to note that compound **5** is the most potent NAM, which might be related to its ability to keep W773^{6.48a.50c} outside the bundle (see below).

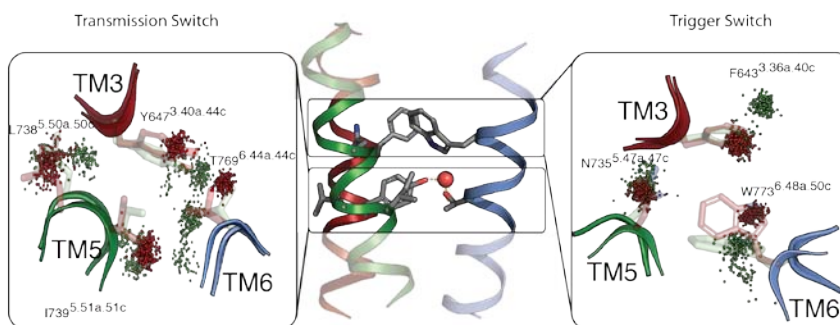


Figure 7 . Comparison of the mechanism of negative and positive allosteric modulation at the mGlu₂ receptor. Relative position of the ‘trigger switch’ amino acids involved in the initial agonist-induced structural changes on the receptor responsible for the rearrangement of the ‘transmission switch’ amino acids that finally lead to receptor activation. Distribution of the positions of C Z atom of F643^{3.36a.40c}, the ND2 atom of N735^{5.47a.47c}, and the NE1 atom of W773^{6.48a.50c} (right panel), and the OH atom of Y647^{3.40a.44c}, CD2 atom of L738^{5.50a.50c}, CG2 atom of I739^{5.51a.51c}, and OG1 atom of T769^{6.44a.46c} (left panel), which are the amino acids of mGlu₂ receptor at the homologous positions of the ‘trigger switch’ and ‘transmission switch’ of class A, respectively, during MD simulations of the “active-like” model of mGlu₂ receptor in complex with Gi and PAM **1** (in green) and of the “inactive” model of mGlu₂ receptor in complex with NAM **6** (in red). Evenly spaced snapshots extracted from the 1 μs of unbiased MD simulations are depicted.

As mentioned above, the PAMs form a crucial hydrogen bond interaction with N735^{5.47a.47c} in TM5. PAMs without this HBA feature lose the ability to allosterically modulate the receptor, and the N735^{5.47a.47c}D mutation makes the compounds inactive. The hydrogen bond between N735^{5.47a.47c} and the PAMs **1-3** remains

constant throughout all the MD simulations, with only small fluctuations to accommodate the part of the molecule that points to the extracellular domain (Supplementary Figure S3). Accordingly, the side chain of N735^{5.47a.47c} stays in a similar position (Figure 6B) during the simulations of PAMs **1-3**. Moreover, in all simulations of PAMs, W773^{6.48a.50c}, which was initially modeled in an outward position pointing into the membrane, modifies its conformation to move inside the receptor (Figure 6B). This new conformation, which has been observed in the mGlu₅-StaR crystal structures,⁵⁴ reduces the volume of the bottom of the receptor cavity. Accordingly, PAMs with longer substituents pointing inside the receptor (EX pharmacophoric feature) ultimately lose their activation activity. In general all PAM molecules present small and flexible substituents that correspond to the HYD1 pharmacophoric feature. We observe in the simulations of PAMs **1-3** that HYD1 interacts with W773^{6.48a.50c}, bringing its side chain inside the receptor (*gauche+* conformation, Figure 6D). In the case of the NAMs we see this behavior for compounds **4** and **6** (see above). Clearly, the W773^{6.48a.50c}A mutation affects both PAM and NAM activity.

In order to compare the mechanism of the negative and positive allosteric modulation at the mGlu₂ receptor in more detail, we plot the movement of the amino acids in the first activation ‘trigger switch’ (F643^{3.36a.40c}, N735^{5.47a.47c} and W773^{6.48a.50c}) and the ‘transmission switch’ (Y647^{3.40a.44c}, L738^{5.50a.50c}, I739^{5.51a.51c} and T769^{6.44a.46c}) during the MD simulations in the presence of **1** and **6**, chosen as representative PAM and NAM ligands (Figure 7). The hydrogen bond interaction of the carbonyl group of PAM **1** with N735^{5.47a.47c}, which is absent in NAM compounds, moves N735^{5.47a.47c} toward TM3, whereas F643^{3.36a.40c} that interacts with both NAM and PAM compounds (Tables 2 and 3), moves toward TM7 in the PAM-bound simulation. This hydrogen bond interaction between PAMs and N735^{5.47a.47c} resembles the interaction between agonists and TM5 in class A, but how does it affect the ‘transmission switch’? Relative to the inactive simulation (NAM **6**), in the active simulation (PAM **1**) there is an inward movement of TM5 (at L738^{5.50a.50c}) toward TM3, relocation of Y647^{3.40a.44c} toward TM6, and finally reposition of the side chain of T769^{6.44a.46c}. This includes a conformational change of the side chain of T769^{6.44a.46c} from the *gauche+* to the *gauche-* or *trans* conformer, Figures 6G-H, which was not observed for NAMs (as commented

above). A conformational change of this type has been shown to cause bend in α -helices⁶⁷ which in this case would facilitate the outward movement of TM6 for receptor activation. Furthermore, this threonine is four amino acids below the tryptophan in the FTMYTTCI*WLAF sequence in TM6 which is conserved for all mGlu receptors.

Figure 7 also shows a water molecule involved in a network of interactions with T769^{6.44a.46c} and Y647^{3.40a.44c} in the 'transmission switch'. This water molecule was observed to enter from bulk solvent in all the simulations despite not being included in the initial system. Via hydrogen bonds with the side chains of T769^{6.44a.46c} and Y647^{3.40a.44c} it helps to stabilize a specific orientation of these amino acids. In the experimental X-ray crystal structure of the mGlu₅ receptor a water molecule was seen in a similar position and was proposed as a key element involved in ligand pharmacology.⁴⁹ The prevalence of the water molecule in the simulations suggests it is a robust feature, and together with the understanding that subtly different interactions with mGlu₅ NAM ligands can reverse their functional effects, supports our conclusions on the important role of these amino acids and hydrogen bonding network for receptor activation.

3.9.4. Conclusions

Here, we addressed the mechanism of positive versus negative allosteric modulation at class C GPCRs, using the mGlu₂ receptor as a model. Our study proposes that modulation of class C GPCRs by either positive (PAMs) or negative (NAMs) allosteric modulators involves rearrangement of homologous ‘switches’ as (in)activation of class A by either orthosteric antagonists or agonists. We have identified an activation ‘trigger switch’ that is rearranged by PAM binding and a ‘transmission switch’ that is not directly involved in ligand interactions but links the binding site with the outward movement of TM6 for receptor activation and G protein binding. These combined experimental and computational results strongly support that despite the low degree of sequence similarity between classes A and C of GPCRs, the two families share conserved elements in their mechanisms of receptor activation. This work will help with the discovery of new generations of mGlu receptor allosteric modulator drugs and the understanding of class C GPCR allosteric modulation.

3.9.5. Methods

Plasmids, cell transfection and cell culture. cDNA constructs encoding human non-mutated and mutated mGlu₂ receptors were synthesized by GeneArt® (Life Technologies), subcloned to the mammalian expression vector p cDNA3.1(+) (Life Technologies) and amplified through *E. coli* transformation. CHO-K1 cells were used for transient transfection. 24 hours prior to transfection, cells were seeded at high density (20,000 cells/cm²) into 14 cm Ø plates. Transfections were performed using lipofectamine LTX reagent (Life Technologies). CHO-K1 cells expressing mutated and non-mutated mGlu₂ receptors were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS, penicillin, streptomycin, pyruvic acid and L-Glutamine. Stably expressing hmGlu₂-CHO cells were also used as internal control. Cells were kept in a humidified atmosphere at 37°C and 5% CO₂ and were subcultured twice weekly.

Membrane Preparation. Cells were detached from the plates by scraping into 50 mM Tris-HCl buffer, pH 7.4 and subsequently centrifuged for 10 min at 16,000 rpm in a Sorvall 5C Plus S34

centrifuge at 4 °C. Pellets were resuspended in ice-cold 5 mM hypotonic Tris-HCl, pH 7.4 and homogenized using an Ultra Turrax homogenizer (IKA-Werke GmbH & Co.KG, Staufen, Germany) at 24,000 rpm. Homogenates were centrifuged at 18,000 rpm for 20 min at 4°C. Remaining pellets were suspended in 50 mM Tris-HCl pH 7.4 and the homogenization step was repeated. Aliquots were stored at -80°C. Protein concentrations were determined using the Bradford method (Bio-Rad, Hercules, CA, USA) using BSA as a standard.

[³⁵S]GTPγS Binding Assay. Membranes were thawed and homogenized using an Ultra Turrax homogenizer at 24,000 rpm. Samples were diluted in ice-cold assay buffer (20 mM HEPES pH 7.4, 100 mM NaCl, 3 mM MgCl₂, and 10 μM GDP and 14.3 μg/mL saponin). DMSO concentrations were ≤1%. Assay mixtures containing a variable concentration of test compound and 10 μg membrane protein were pre-incubated with buffer (to detect agonist effects) or an EC₂₀ or EC₈₀-equivalent concentration of glutamate (to detect PAM or NAM effects respectively). After 30 minutes of incubation at 30 °C, 0.1 nM [³⁵S]GTPγS was added. The reaction was stopped after another 30 minute incubation at 30 °C by rapid

filtration through a 96-well GF/B filterplate (PerkinElmer) on a PerkinElmer filtermate harvester. Plates were washed three times with ice-cold wash buffer (10 mM NaH₂PO₄/10 mM Na₂HPO₄, pH 7.4) and dried overnight. Filter-bound radioactivity was counted in a Topcount microplate scintillation and luminescence counter (PerkinElmer).

[³H]JNJ-46281222 Binding Assay. Membranes were allowed to thaw and subsequently homogenization was performed using an Ultra Turrax homogenizer at 24,000 rpm. Samples were diluted in ice-cold assay buffer (50 mM Tris-HCl pH 7.4, 10 mM MgCl₂ and 2 mM CaCl₂). Binding assays were performed in glass tubes in a total volume of 500 μl, containing variable concentrations of test compound, 75 μg membrane protein of the stably expressed hmGlu2 and 4 nM [³H]JNJ-46281222. Nonspecific binding was determined using 10 μM JNJ-42341806. The reaction was stopped after incubation for 1 hour at room temperature by rapid filtration over pre-coated (PEI 0.1%, Sigma-Aldrich) GF/C filters through a Brandel harvester 96 (Brandel, Gaithersburg, MD, USA). Filters were washed three times with ice-cold wash buffer (50 mM Tris-HCl pH 7.4). Filter-bound radioactivity was determined using liquid

scintillation spectrometry on a Tri-Carb 2810TR counter (PerkinElmer).

Data analysis. Data analyses were performed using Prism 4.02 (GraphPad software, La Jolla, CA, USA). For the datasets of [³H]JNJ-46281222 binding experiments, pIC₅₀ values were obtained using non-linear regression curve fitting into a sigmoidal concentration-response curve using the equation: $Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / \{1 + 10^{(X - \text{LogIC}_{50})}\}$. pK_i values were obtained from pIC₅₀ values using the Cheng-Prusoff⁶⁸ equation. Concentration-response curves obtained in [³⁵S]GTPγS binding experiments were fitted using non-linear regression curve fitting into a sigmoidal concentration-response curve using the equation: $Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / \{1 + 10^{[(\text{LogEC}_{50} - X) \times \text{Hill Slope}]}\}$. Statistical analysis was performed if indicated, using one-way ANOVA with Dunnett's post-test.

Preparing datasets for pharmacophore building. Known mGlu₂ receptor PAM and NAM compounds were retrieved from ChEMBL⁶⁹ searching for keywords 'GRM2' and 'mGlu2'. Molecules were retained only if they had concentration response

bioactivity less than 1 μ M (all having ‘ChEMBL Confidence Score’ 8 and 9). Orthosteric ligands containing amino acid substructures were removed. By this approach 296 PAMs were identified. For NAMs, only 17 active molecules were identified in ChEMBL which matched the activity type “antagonist” or “negative allosteric”. Due to the low number, the NAMs dataset was augmented by mining patents. Hence a further 275 NAMs were retrieved from 12 patents (WO2007110337, WO2005040171, WO2008128889, WO2008119689, WO2006099972, WO2006084634, WO2005123738, WO2003066623, WO2002023665, WO2002083652, WO2001129012, US20070072879). The complete NAM active dataset contained 289 unique mGlu₂ receptor active NAMs.

Pharmacophore elucidation requires known inactive molecules that are often absent in public bioactivity databases. Hence these were taken from Janssen in-house mGlu₂ receptor PAM and NAM high throughput screening (HTS) data. Inactives were retained in a similar MW range to the PAMs (233 to 515 Dalton) and NAMs (303 to 658 Dalton). A subset was selected to initially derive the pharmacophore in an automated manner. Hence, in the case of PAMs 86 molecules were selected whereas for NAMs 82 were

chosen. These were combined with two randomly selected sets of 235 molecules shown to be inactive in either the Janssen PAM or NAM HTS.

PAM Pharmacophore. Pharmacophore elucidation was performed on the smaller training set (86 active and 235 inactive PAMs). The actives were structurally similar to PAMs **1**, **2** and **3**. Molecules were prepared using MOE.⁷⁰ First the wash function was employed to assign protonation and tautomeric states as applicable. The ph4elucidate tool in MOE was used with default settings and with 3D conformers generated by the conformational import approach. The automated elucidation delivered a satisfactory alignment with the placement of four features. The pharmacophore with best statistical retrieval of actives (85 molecules, 99 % hit) and inactives (108 molecules, 46 % hit) was chosen then augmented by the inclusion of additional features. The final PAM pharmacophore hypothesis was submitted to further statistical testing by searching on the full test set of actives (296 molecules) and inactives (2791 molecules). From this testing 121 of the 296 (41 %) of the PAM actives were hit whereas only 204 of the 2791 (7 %) of the HTS inactives were hit. The 121 examples that were hit by the final

pharmacophore were analogues of the specific chemical series of interest in this study. The pharmacophore did not hit all the PAM chemical series. This is expected as it is unlikely for a single pharmacophore to cover diverse chemical series.

NAM Pharmacophore. A smaller training set (82 active and 235 inactive NAMs) was used as input for the ph4elucidate tool in MOE with all settings the same as described for PAMs. NAM actives were more structurally dissimilar including benzodiazapinones, and bicyclic pyrazolopyrimidines. The automated elucidation delivered one satisfactory alignment which showed good structural and feature overlap of actives from more than one chemical series. The five feature pharmacophore had good statistical behavior hitting 70 of the 82 (85 %) active NAMs but only 41 of the 235 (17 %) inactive molecules. As in the case of the PAMs, the initial pharmacophore was augmented by incorporating additional features based on knowledge of the SAR. The final 11 feature pharmacophore was subjected to further statistical testing on the entire set of active (289) and inactive (2800) NAMs. The requirement to hit all 11 features was too restrictive, hence partial matching was enabled to permit molecules to hit some of the 11

features. Good statistical performance was seen with partial matching set to 9 of the 11 features, which hit 59 % (171) of the actives and only 1 % (28) of the inactives. Hence both PAM and NAM pharmacophores were able to discriminate known actives and inactives and captured details of the SAR.

Computational models of the mGlu₂ receptor. Two structural models (inactive and “active-like” states) of the 7-TM domain of human mGlu₂ receptor (Uniprot code Q14416) were built using a combination of structural templates. The crystal structure of inactive mGlu₅ receptor (Protein Data Bank (PDB)⁷¹ code 4OO9) was used for the construction of inactive mGlu₂ (these receptors share 51% of sequence identity at the 7-TM domain). Due to the absence of ECL2 in the crystal structure of the highly homologous mGlu₅ receptor, ECL2 of mGlu₂ was modeled using the crystal structure of mGlu₁ receptor (PDB 4OR2). To study mGlu₂ activation by PAMs, we needed an “active-like” model. This was generated from the crystal structure of the β_2 AR-Gs complex (PDB ID 3SN6), by changing the conformation of TM6 of mGlu₂ receptor N757^{6.32a.34c}-M766^{6.41a.43} for the active conformation of β_2 - 6.32-6.41. This single replacement opens the intracellular cavity required for the binding

of the C-terminal $\alpha 5$ helix of the G-protein. It has been shown that an agonist alone in class A (or analogously a PAM in class C) is not capable of stabilizing the fully active conformation of the receptor in the absence of the G protein,⁷² hence, this “active-like” model of mGlu₂ receptor includes G_i. It is important to remark that the extracellular part, including receptor side chain conformations, of the “active-like” model of mGlu₂ receptor is comparable to the inactive model. The α -helical domain of G_i α was modeled in the “closed” conformation,⁷³ using the crystal structure of [AlF₄]⁻-activated G_i (PDB code 1AGR)⁷⁴. A similar approach has recently been used.^{32,75} These initial models were constructed in MOE, and Maestro⁷⁶ was used for structure preparation. The Protein Preparation⁷⁷ tool was used to fix any missing sidechains/atoms, PROPKA⁷⁸ assigned protonation states, the hydrogen bonding network was optimized, and brief minimization to RMSD 0.5 Å was applied to remove any structural clashes.

Docking of NAMs into the mGlu₂ receptor. Multiple conformers (calculated with the ConfGen⁷⁹ module of Maestro) of the ligands were docked into the inactive models of mGlu₂ receptor

using Glide XP.⁸⁰ The docking grid was centered on the center of mass of the ligand position in the mGlu₁ receptor structure. Sampling was increased for the Glide (Schrodinger LLC, New York, NY, USA) docking by turning on expanded sampling and passing 100 initial poses to post-docking minimization. All other docking parameters were set to the defaults.

Molecular dynamics: Molecular dynamics (MD) simulations were performed with GROMACS v5.0.6.⁸¹ The complexes between NAMs and the inactive conformation of mGlu₂ receptor and PAMs and the “active” conformation of mGlu₂ receptor in complex with Gi were embedded in a pre-equilibrated box (9x9x9 or 10x10x19 nm, respectively) containing a lipid bilayer (205 or 297 POPC molecules) with explicit solvent (~14000 or ~47000 waters) and 0.15 M concentration of Na⁺ and Cl⁻ (~140 or ~490 ions). Each system was energy minimized and subjected to a 5-step MD equilibration (10+5+2+2+2 ns). In the first step the whole system was fixed except hydrogen atoms; in the second step, the protein loops were released from restraints; and in the final three steps the restraints on the ligand and protein atoms were relaxed from 100, 50

to $10 \text{ kJ}\cdot\text{mol}^{-1}\text{nm}^{-2}$, respectively. Unrestrained MD trajectories were produced for $1 \mu\text{s}$ using a 2 fs time step and a total of three independent $1 \mu\text{s}$ simulations were performed for each ligand-receptor system. Constant temperature of 300 K using separate variable rescale thermostats for protein-ligand, lipids, and water plus ions was used. The LINCS algorithm was applied to freeze bond lengths. Lennard-Jones interactions were computed using a 10 \AA cut-off, and the electrostatic interactions were treated using PME with a direct sum cut-off of 10 \AA . The AMBER99SD-ILDN force field⁸² was used for the protein, the parameters described by Berger *et al.*⁸³ for lipids, and the general Amber force field (GAFF) and HF/6-31G*-derived RESP atomic charges for the ligand. This combination of protein and lipid parameters has recently been validated.⁸⁴

3.9.6. Supporting information

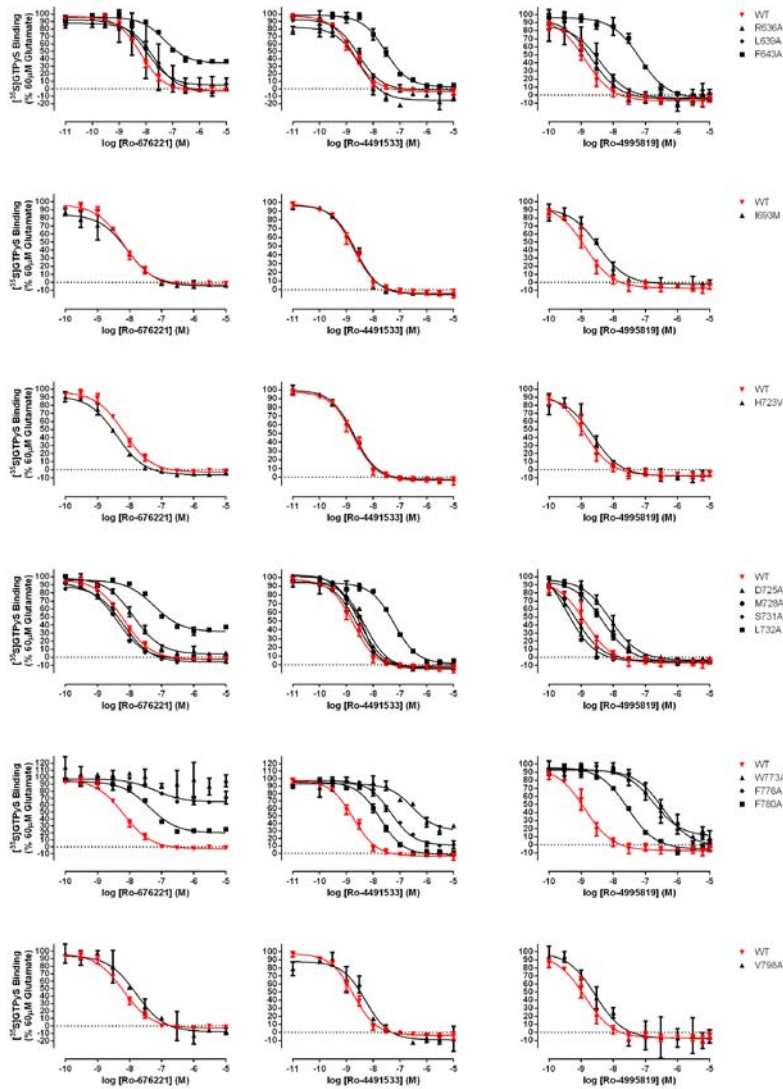


Figure S1. Concentration response curves for each mGlu2 receptor NAM, **4** Ro-676221, **5** Ro-4491533, and **6** Ro-4995819 tested in the WT and mGlu2 receptor mutants.

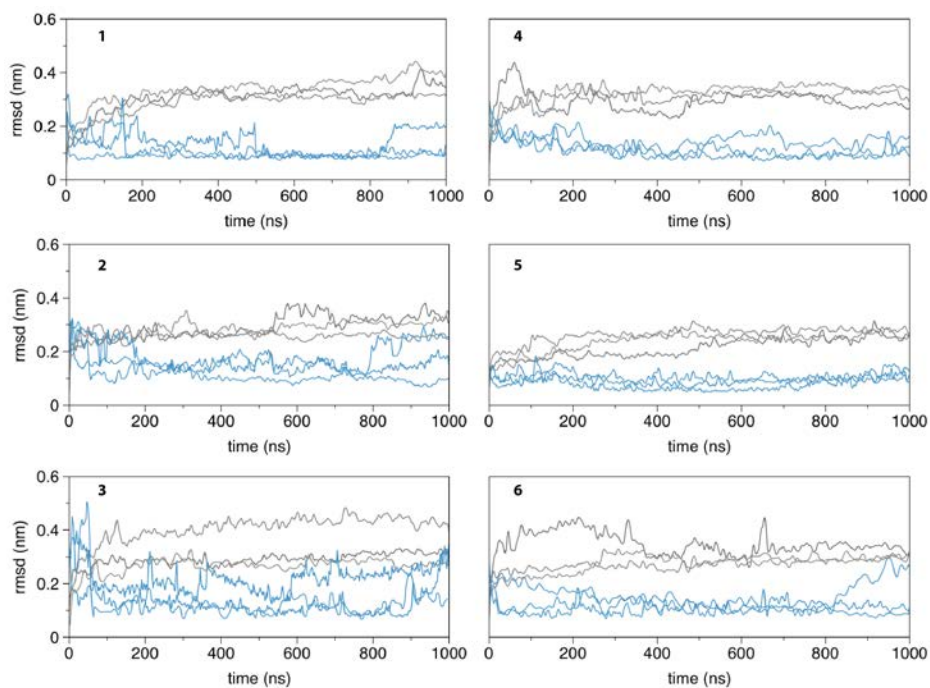


Figure S2. Root mean-square deviation (RMSD) of the receptor backbone atoms (black) and ligand heavy atoms (blue) of the MD simulations (3 replicas of unbiased 1 μ s) of the mGlu2 receptor in complex with Gi and PAMs **1-3** (left column) and of the mGlu2 receptor in complex with NAMs **4-6** (right column).

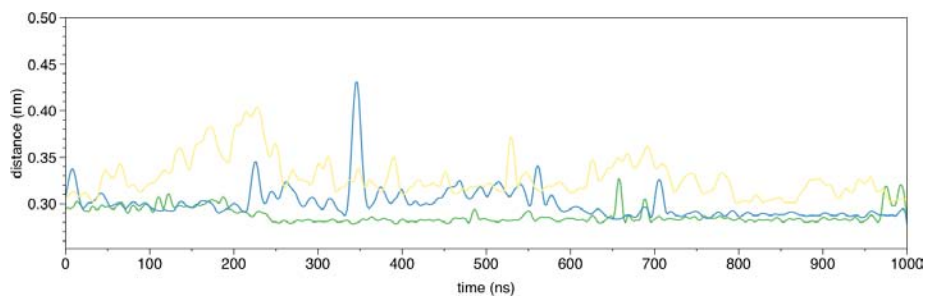


Figure S3. Time-evolution of the hydrogen bond distance between N735^{5.47a.47c} and the carbonyl group of PAMs **1** (green) or **2** (blue) and the nitrogen atom of the triazo ring of PAM **3** (yellow), along the MD trajectories.

Selectivity versus mGlu receptors: PAMs are selective but NAMs inhibit mGlu3 receptors

Molecules **1** to **6** were tested in a functional mGlu receptor assay panel (see Methods and Table 1). All PAMs did not activate any of the other human mGlu receptor subtypes or the rat mGlu6 receptor up to 10 μ M concentration limit. However, they did show weak mGlu2 agonistic activity as described previously for **1**,¹ **2**² and **3**³ which may be due to residual levels of endogenous glutamate. PAMs did not inhibit glutamate-induced signaling at any of the receptors up to 10 μ M concentration. Considering NAMs **4**, **5** and **6**, no activation of any mGlu receptors was seen up to the same concentration limit. Importantly however, NAMs showed antagonistic inhibitory effects at mGlu3 receptors in a comparable range to their activity in the analogous assay at mGlu2R. Hence, NAMs are not selective versus mGlu3 receptors.

Table S1. In vitro selectivity of PAMs (**1** to **3**) and NAMs (**4** to **6**) in an mGlu receptor functional activity panel

Compound	Activation (pIC ₅₀)	Inhibition (pIC ₅₀)
1, BINA	mGlu2 5.5 mGlu1, 3, 4, 5, 6, 7 and 8 all < 5	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5
2, JNJ-40068782	mGlu2 6.2 mGlu1, 3, 4, 5, 6, 7 and 8 all < 5	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5
3, JNJ-46281222	mGlu2 6.7 mGlu1, 3, 4, 5, 6, 7 and 8 all < 5	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5
4, Ro-676221	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5	mGlu2 6.4, mGlu3 6.3 mGlu1, 4, 5, 6, 7 and 8 all < 5
5, Ro-4491533	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5	mGlu2 6.6, mGlu3 6.4 mGlu1, 4, 5, 6, 7 and 8 all < 5
6, Ro-4995819	mGlu1, 2, 3, 4, 5, 6, 7 and 8 all < 5	mGlu2 6.9, mGlu3 6.5 mGlu1, 4, 5, 6, 7 and 8 all < 5

Method.

mGlu receptor panel selectivity assays: Ca^{2+} assays with human mGlu1, 3, 5, 7, or 8 receptor expressing HEK 293 cells were performed as reported in Lavreysen et al. (2013), except for a slight change in the procedure for mGlu5: cells expressing the human mGlu5 receptor were seeded at 40,000 cells/well in MW384. Twenty-four hours after seeding, cells were incubated for 90 min in Ca^{2+} assay kit (Molecular Devices) dissolved in saline PBS supplemented with 5 mmol/L probenecid, pH 7.4 (f.c. 2.5 mmol/L probenecid as loading buffer was added on the cell layer without removal of medium) before measurements. Measurement of $[^{35}S]GTP\gamma S$ binding to membranes from CHO cells expressing the rat mGlu6 receptor and membranes from L929sA cells expressing the human mGlu4 receptor were conducted also as described in Lavreysen et al. 2013.

3.9.7. References

- 1 Siegel, G. J. *Basic neurochemistry : molecular, cellular, and medical aspects*. 7th ed. edn, (Burlington, MA : Elsevier Academic, c2006., 2006).
- 2 Niswender, C. M. & Conn, P. J. Metabotropic glutamate receptors: physiology, pharmacology, and disease. *Annu. Rev. Pharmacol. Toxicol.* **50**, 295-322, doi:10.1146/annurev.pharmtox.011008.145533 (2010).
- 3 Vinson, P. N. & Conn, P. J. Metabotropic glutamate receptors as therapeutic targets for schizophrenia. *Neuropharmacology* **62**, 1461-1472, (2012).
- 4 Kew, J. N. C. & Kemp, J. A. Ionotropic and metabotropic glutamate receptors structure and pharmacology. *Psychopharmacology* **179**, 4-29, (2005).
- 5 Richards, G. *et al.* Distribution and abundance of metabotropic glutamate receptor subtype 2 in rat brain revealed by [3H]LY354740 binding in vitro and quantitative radioautography: Correlation with the sites of synthesis, expression, and agonist stimulation of [35S]GTP γ S binding. *J. Comp. Neurol.* **487**, 15-27, (2005).
- 6 Gu, G. *et al.* Distribution of metabotropic glutamate 2 and 3 receptors in the rat forebrain: Implication in emotional responses and central disinhibition. *Brain Res* **1197**, 47-62, (2008).
- 7 Ghose, S., Gleason, K. A., Potts, B. W., Lewis-Amezcu, K. & Tamminga, C. A. Differential expression of metabotropic glutamate receptor 2 and 3 in schizophrenia: a mechanism for antipsychotic drug action? *Am J Psychiatry* **166**, 812-820, (2009).
- 8 Celanire, S. *et al.* Novel metabotropic glutamate receptor 2/3 antagonists and their therapeutic applications: a patent review (2005 – present). *Expert Opin. Ther. Pat.* **25**, 69-90, (2015).
- 9 *Addex moves mGluR2 antagonist into clinical trials for Alzheimer's disease available online:* <http://www.Bioportfolio.Com/news/article/120812/addex-makes-plans-to-move-mglur2-antagonist-into-clinical-trials-for-alzheimer.html>.

- 10 Monn, J. A. *et al.* Design, synthesis, and pharmacological characterization of (+)-2-aminobicyclo[3.1.0]hexane-2,6-dicarboxylic Acid (LY354740): A potent, selective, and orally active group 2 metabotropic glutamate receptor agonist possessing anticonvulsant and anxiolytic Properties. *J. Med. Chem.* **40**, 528-537 (1997).
- 11 Fell, M. J., McKinzie, D. L., Monn, J. A. & Svensson, K. A. Group I metabotropic glutamate receptor agonists and positive allosteric modulators as novel treatments for schizophrenia. *Neuropharmacology* **62**, 1473-1483, (2012).
- 12 Patil, S. T. *et al.* Activation of mGlu2/3 receptors as a new approach to treat schizophrenia: a randomized Phase 2 clinical trial. *Nat Med* **13**, 1102-1107, (2007).
- 13 Dunayevich, E. *et al.* Efficacy and tolerability of an mGlu2/3 agonist in the treatment of generalized anxiety disorder. *Neuropsychopharmacology* **33**, 1603-1610 (2007).
- 14 Conn, P. J., Christopoulos, A. & Lindsley, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat. Rev. Drug. Discov.* **8**, 41-54, (2009).
- 15 Christopoulos, A. *et al.* International Union of Basic and Clinical Pharmacology. X. C. mu ltisite pharmacology: recommendations for the nomenclature of receptor allosterism and allosteric ligands. *Pharmacol. Rev.* **66**, 918-947, (2014).
- 16 Trabanco, A. A., Cid, J. M., Lavreysen, H., Macdonald, G. J. & Tresadern, G. Progress in the development of positive allosteric modulators of the metabotropic glutamate receptor 2. *Curr. Med. Chem.* **18**, 47-68, (2011).
- 17 Trabanco, A. A. & Cid, J. M. mGluR2 positive allosteric modulators: a patent review (2009 – present). *Expert Opin. Ther. Pat.* **23**, 629-647, (2013).
- 18 Cid, J. M. *et al.* Discovery of 1,5-disubstituted pyridones: A new class of positive allosteric modulators of the metabotropic glutamate 2 receptor. *ACS Chem. Neurosci.* **1**, 788-795, (2010).
- 19 Cid, J. M. *et al.* Discovery of 1,4-disubstituted 3-cyano-2-pyridones: A new class of positive allosteric modulators of the metabotropic glutamate 2 receptor. *J. Med. Chem.* **55**, 2388-2405, (2012).
- 20 Fell, M. J. *et al.* N-(4-((2-(trifluoromethyl)-3-hydroxy-4-(isobutyryl)phenoxy)methyl)benzyl)-1-methyl-1H-

- imidazole-4-carboxamide (THIC), a novel metabotropic glutamate 2 potentiator with potential anxiolytic/antidepressant properties: In vivo profiling suggests a link between behavioral and central nervous system neurochemical changes. *J. Pharmacol. Exp. Ther.* **336**, 165-177 (2011).
- 21 Johnson, P. L. *et al.* Group II metabotropic glutamate receptor type 2 allosteric potentiators prevent sodium lactate-induced panic-like response in panic-vulnerable rats. *J. Psychopharmacol. (Oxford, England)* **27**, 152-161, (2013).
- 22 Pinkerton, A. B. *et al.* Phenyl-tetrazolyl acetophenones: Discovery of positive allosteric potentiators for the metabotropic glutamate 2 receptor. *J. Med. Chem.* **47**, 4595-4599, (2004).
- 23 Tresadern, G. *et al.* Scaffold hopping from pyridones to imidazo[1,2-a]pyridines. New positive allosteric modulators of metabotropic glutamate 2 receptor. *Bioorg. Med. Chem. Lett.* **20**, 175-179, (2010).
- 24 Cid, J. M. *et al.* Discovery of 3-cyclopropylmethyl-7-(4-phenylpiperidin-1-yl)-8-trifluoromethyl[1,2,4]triazolo[4,3-a]pyridine (JNJ-42153605): A positive allosteric modulator of the metabotropic glutamate 2 receptor. *J. Med. Chem.* **55**, 8770-8789, (2012).
- 25 Trabanco, A. A. *et al.* Imidazo[1,2-a]pyridines: Orally active positive allosteric modulators of the metabotropic glutamate 2 receptor. *J. Med. Chem.* **55**, 2688-2701, (2012).
- 26 Lavreysen, H. *et al.* Preclinical evaluation of the antipsychotic potential of the mGlu2-positive allosteric modulator JNJ-40411813. *Pharmacol. Res. Perspect.* **3**, e00097, (2015).
- 27 AstraZeneca.
<http://www.astrazeneca.com/Investors/financial-information/Financial-results/2010-Financial-results>.
- 28 *Study to explore the Safety, Tolerability and Potential clinical efficacy of JNJ-40411813 in schizophrenic patients.*,
<http://clinicaltrials.gov/show/NCT01323205>.
- 29 Cid, J. M. *et al.* Discovery of 1-Butyl-3-chloro-4-(4-phenyl-1-piperidinyl)-(1H)-pyridone (JNJ-40411813): A novel positive allosteric modulator of the metabotropic glutamate 2 receptor. *J. Med. Chem.* **57**, 6495-6512, (2014).

- 30 Galici, R. *et al.* Biphenyl-indanone A, a positive allosteric modulator of the metabotropic glutamate receptor subtype 2, has anxiolytic- and anxiolytic-like effects in mice. *J. Pharmacol. Exp. Ther.* **318**, 173-185 (2006).
- 31 Lavreysen, H. *et al.* Pharmacological characterization of JNJ-40068782, a new potent, selective, and systemically active positive allosteric modulator of the mGlu2 receptor and its radioligand [³H]JNJ-40068782. *J. Pharmacol. Exp. Ther.* **346**, 514-527 (2013).
- 32 Doornbos, M. L. *et al.* Molecular mechanism of positive allosteric modulation of the metabotropic glutamate receptor 2 by JNJ-46281222. *Br. J. Pharmacol.* **173**, 588-600, (2016).
- 33 Nicholls, R. E. *et al.* mGluR2 acts through inhibitory G α subunits to regulate transmission and long-term plasticity at hippocampal mossy fiber-CA3 synapses. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6380-6385 (2006).
- 34 Campo, B. *et al.* Characterization of an mGluR2/3 negative allosteric modulator in rodent models of depression. *J. Neurogenet.* **25**, 152-166, (2011).
- 35 Ahnaou, A., Ver Donck, L. & Drinkenburg, W. H. I. M. Blockade of the metabotropic glutamate (mGluR2) modulates arousal through vigilance states transitions: Evidence from sleep-wake EEG in rodents. *Behav. Brain Res.* **270**, 56-67, (2014).
- 36 Goeldner, C. *et al.* Cognitive impairment in major depression and the mGlu2 receptor as a therapeutic target. *Neuropharmacology* **64**, 337-346, (2013).
- 37 Study, A. *A Study of RO4995819 in Patients With Major Depressive Disorder And Inadequate Response to Ongoing Antidepressant Treatment*, <<https://www.clinicaltrials.gov/ct2/show/NCT01457677>> (
- 38 Gatti, M. S., Saxe, M., Wichmann, J. & Woltering, T. F. mGlu2/3 antagonists for the treatment of autistic disorders. WO Patent Application WO064028 filed 21 Oct. 2013, and published 1 May 2014.
- 39 Kniazeff, J., Prézeau, L., Rondard, P., Pin, J.-P. & Goudet, C. Dimers and beyond: The functional puzzles of class C GPCRs. *Pharmacol. Ther.* **130**, 9-25, (2011).

- 40 Huang, S. *et al.* Interdomain movements in metabotropic glutamate receptor activation. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15480-15485 (2011).
- 41 Xue, L. *et al.* Major ligand-induced rearrangement of the heptahelical domain interface in a GPCR dimer. *Nat Chem Biol* **11**, 134-140, (2015).
- 42 Farinha, A. *et al.* Molecular determinants of positive allosteric modulation of the human metabotropic glutamate receptor 2. *Br. J. Pharmacol.* **172**, 2383-2396, (2015).
- 43 Hemstapat, K. *et al.* A novel family of potent negative allosteric modulators of group II metabotropic glutamate receptors. *J. Pharmacol. Exp. Ther.* **322**, 254-264 (2007).
- 44 Lundström, L. *et al.* Structural determinants of allosteric antagonism at metabotropic glutamate receptor 2: mechanistic studies with new potent negative allosteric modulators. *Br. J. Pharmacol.* **164**, 521-537, (2011).
- 45 Rowe, B. A. *et al.* Transposition of three amino acids transforms the human metabotropic glutamate receptor (mGluR)-3 positive allosteric modulation site to mGluR2, and an additional characterization of the mGluR2-positive allosteric modulation site. *J. Pharmacol. Exp. Ther.* **326**, 240-251 (2008).
- 46 Schaffhauser, H. *et al.* Pharmacological characterization and identification of amino acids involved in the positive modulation of metabotropic glutamate receptor subtype 2. *Mol. Pharmacol.* **64**, 798-810 (2003).
- 47 El Moustaine, D. *et al.* Distinct roles of metabotropic glutamate receptor dimerization in agonist activation and G-protein coupling. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16342-16347 (2012).
- 48 Goudet, C. *et al.* Asymmetric functioning of dimeric metabotropic glutamate receptors disclosed by positive allosteric modulators. *J. Biol. Chem.* **280**, 24380-24385 (2005).
- 49 Dore, A. S. *et al.* Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* **511**, 557-562, (2014).
- 50 Wu, H. *et al.* Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* **344**, 58-64, (2014).

- 51 Isberg, V. *et al.* Generic GPCR residue numbers - aligning topology maps minding the gaps. *Trends Pharmacol. Sci.* **36**, 22-31, (2015).
- 52 Tresadern, G., Cid, J.-M. & Trabanco, A. A. QSAR design of triazolopyridine mGlu2 receptor positive allosteric modulators. *J. Mol. Graph. Model.* **53**, 82-91, (2014).
- 53 Gregory, K. J. *et al.* Identification of specific ligand-receptor interactions that govern binding and cooperativity of diverse modulators to a common metabotropic glutamate receptor 5 allosteric site. *ACS Chem. Neurosci.* **5**, 282-295, (2014).
- 54 Christopher, J. A. *et al.* Fragment and structure-based drug discovery for a class C GPCR: discovery of the mGlu5 negative allosteric modulator HTL14242 (3-chloro-5-[6-(5-fluoropyridin-2-yl)pyrimidin-4-yl]benzotrile). *J. Med. Chem.* **58**, 6653-6664, (2015).
- 55 Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor. *Nature* **469**, 175-180, (2011).
- 56 Deupi, X. & Kobilka, B. K. Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. *Physiology (Bethesda)* **25**, 293-303, (2010).
- 57 Dror, R. O. *et al.* Identification of two distinct in active conformations of the beta2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. USA* **106**, 4689-4694, (2009).
- 58 Stanley, N., Pardo, L. & Fabritiis, G. D. The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* **6**, 22639, (2016).
- 59 Codomí, A. *et al.* Functional elements of the gastric inhibitory polypeptide receptor: Comparison between secretin- and rhodopsin-like G protein-coupled receptors. *Biochem. Pharmacol.* **96**, 237-246, (2015).
- 60 Spyridaki, K. *et al.* Structural-functional analysis of the third transmembrane domain of the corticotropin-releasing factor type 1 receptor: role in activation and allosteric antagonism. *J. Biol. Chem.* **289**, 18966-18977, (2014).
- 61 Venkatakrisnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185-194, (2013).
- 62 Sansuk, K. *et al.* A Structural insight into the reorientation of transmembrane domains 3 and 5 during family A G

- protein-coupled receptor activation. *Mol. Pharmacol.* **79**, 262-269, (2011).
- 63 Warne, T. *et al.* The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. *Nature* **469**, 241-244, (2011).
- 64 Choe, H. W. *et al.* Crystal structure of metarhodopsin II. *Nature* **471**, 651-655, (2011).
- 65 Xu, F. *et al.* Structure of a nan agonist-bound human A_{2A} adenosine receptor. *Science* **332**, 322-327, (2011).
- 66 Pellissier, L. *et al.* Conformational toggleswitches implicated in basal constitutive and agonist-induced activated states of 5-HT₄ receptors. *Mol. Pharmacol.* **75**, 982-990 (2009).
- 67 Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. J. & Pardo, L. Serine and Threonine Residues Bend α -Helices in the $\chi_1 = g^-$ Conformation. *Biophys. J.* **79**, 2754-2760, (2000).
- 68 Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 percent inhibition (I₅₀) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099-3108, (1973).
- 69 Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42**, (2014).
- 70 MOE v. v2014.9 (Chemical Computing Group Inc <https://www.chemcomp.com/>).
- 71 Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235-242 (2000).
- 72 Nygaard, R. *et al.* The dynamic process of beta(2)-adrenergic receptor activation. *Cell* **152**, 532-542, (2013).
- 73 Chung, K. Y. *et al.* Conformational changes in the G protein G_s induced by the beta2 adrenergic receptor. *Nature* **477**, 611-615, (2011).
- 74 Tesmer, J. J. G., Berman, D. M., Gilman, A. G. & Sprang, S. R. Structure of RGS4 bound to AF4-activated Gial: Stabilization of the transition state for GTP hydrolysis. *Cell* **89**, 251-261 (1997).
- 75 Navarro, G. *et al.* Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with Gi and Gs. *BMC Biol.* **14**, 1-12, (2016).
- 76 Schrodinger LLC (120 West 45th Street 17th Floor, Tower 45 New York, NY 10036-4041, 2015).

- 77 Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27**, 221-234, (2013).
- 78 Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pKa values. *J. Chem. Theory Comput.* **7**, 2284-2295, (2011).
- 79 Watts, K. S. *et al.* ConfGen: A conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **50**, 534-546, (2010).
- 80 Friesner, R. A. *et al.* Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **49**, 6177-6196, (2006).
- 81 Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19-25, (2015).
- 82 Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950-1958, (2010).
- 83 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.* **72**, 2002-2013 (1997).
- 84 Cordero, A., Caltabiano, G. & Pardo, L. Membrane protein simulations using AMBER force field and Berger lipid parameters. *J. Chem. Theory Comput.* **8**, 948-958, (2012).

4. Application of Free-energy perturbations

4.1. Alzheimer's disease

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by cognitive and behavioral symptoms that progressively impair function. It is the most common form of dementia contributing to 60-70% of cases.¹ The cognitive symptoms include: memory loss, disorientation, confusion and problems with reasoning and thinking. Behavioral symptoms include: agitation, anxiety, delusions, depression, hallucinations and insomnia.²

The most common form of AD is sporadic, or late onset AD (typically > 65 years), and has genetic and environmental risk factors. The only genetic link is to the $\epsilon 4$ allele of the apolipoprotein E.³ Environmental risk factors proposed for AD include: advanced age (the greatest risk factor for AD),⁴ family history,⁵ traumatic brain injury,^{6,7} cardiovascular diseases⁸ and diet.⁹ In contrast, early onset familial AD, has strong genetic links. More than 160 highly penetrant but rare mutations have been described in three genes: amyloid precursor protein (APP),¹⁰ Presenilin 1 (PSN1) and Presenilin 2 (PSN2) genes.¹¹ All these genes encode proteins that are involved in the amyloid cascade and can lead to amyloid β (A β)

aggregation and therefore the expected onset of the disease, described in more detail below.

4.2. AD hypotheses

The causes of AD are unknown. Several hypotheses exist which include the Tau, cholinergic, glutamate or amyloid pathways. Briefly, the hyperphosphorylation of Tau protein promotes formation of aggregates known as neurofibrillary tangles, which are toxic deposits seen in AD brain.¹² The cholinergic system is responsible for cognitive symptoms in dementia and disturbances have long been linked with AD patients.¹³ Ultimately dysfunction leads to fewer cholinergic neurons and lower acetylcholine levels in the brain impairing memory function.¹⁴ Acetylcholinesterase inhibitors are used as symptomatic treatment for AD and can slow disease progression, but their effects are modest.^{15,16} Glutamate is the main excitatory neurotransmitter in the CNS and mediates critical synaptic transmission for the normal functioning of the nervous system.^{17,18} Glutamate-mediated excitotoxicity via over-activation of N-methyl-D-aspartate (NMDA) receptors leading to excessive Ca^{2+} influx can cause neuronal cell death and is linked

with AD.^{17,18} Finally, the amyloid hypothesis has dominated AD research for the past twenty years.^{19,20} Amyloid β ($A\beta$) peptides aggregate to form toxic protein plaques which lead to synaptic failure, neurodegeneration and cognitive dysfunction. The $A\beta$ peptides are formed through sequential proteolytic cleavage of the amyloid precursor protein (APP), one of the most abundant proteins in the CNS.²¹ $A\beta$ peptides can range from 36- to 43-amino acids in length with some preferring to form aggregates over others.²² A shift in the equilibria towards the amyloidogenic over the non-amyloidogenic APP processing pathway predisposes for AD pathology.²³

4.3. $A\beta$ production and β -secretase

In a healthy brain APP is cleaved first by α -secretase generating a secreted APP derivative ($sAPP\alpha$) and a shorter transmembrane C-terminal fragment of 83 amino acids (C83) (see Figure 1). C83 is subsequently cleaved by the transmembrane γ -secretase complex forming a non-toxic 3 kDa peptide (p3).²⁴⁻²⁶ In contrast to $A\beta$ peptides, $sAPP\alpha$ plays a key role in neuronal survival and plasticity.²⁷ In the amyloidogenic brain, cleavage of APP is

mediated by the β -site APP cleaving enzyme 1 (β -secretase, BACE1) instead of α -secretase. This first processing generates a shorter secreted amino terminal APP derivative called sAPP β along with a membrane-inserted C-terminal fragment of 99 amino acids, C99. This C-terminal peptide is then cleaved by γ -secretase producing A β peptides which are liberated into the extracellular space.²⁵

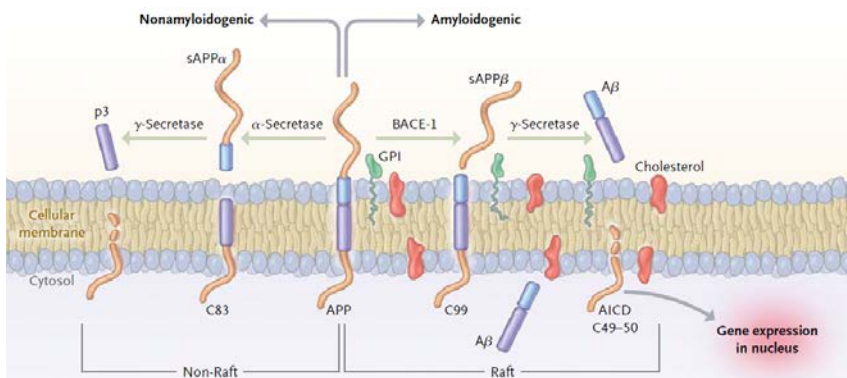


Figure 1. Schematic representation of the amyloid precursor protein processing: amyloidogenic and non-amyloidogenic pathways.

Hence γ -secretase yields A β species of different lengths but A β 40 and A β 42 represent the majority of the species observed in vivo. The longer A β 42 is the more toxic due to higher aggregation potential.²⁸ Therefore, inhibition of the amyloidogenic pathway may be beneficial for AD. Pharmaceutical companies are actively

searching for small molecules that can decrease A β production by affecting the three main enzymes involved in the processing of APP: α -, β - and γ -secretases.

4.4. BACE1 as a drug discovery target

Since its discovery in 1999, BACE1 has become the prime target in AD and the major focus of drug discovery efforts in the pharmaceutical industry. Multiple groups independently discovered the β -secretase enzyme and named it β -site APP cleaving enzyme (BACE), Asp2 or memapsin 2.²⁹⁻³³ BACE1 is a type I monomeric transmembrane aspartic protease related to the pepsin family containing 501 amino acids. Its catalytic domain contains two spatially adjacent aspartate residues that are ~200 residues apart in sequence (Asp32 and Asp228) located between two lobes: the N- and C-terminal halves (Figure 2).³⁴

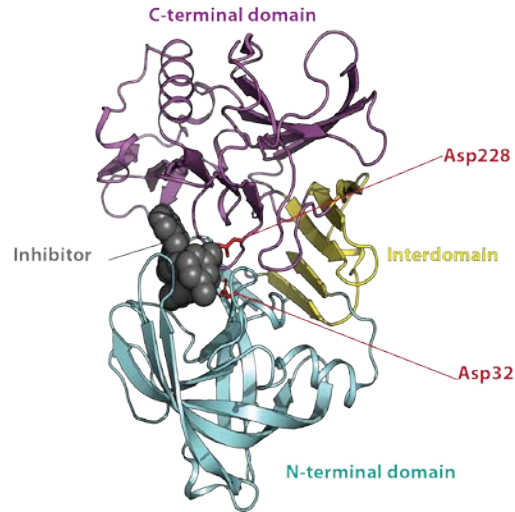


Figure 2. Structure of BACE1 catalytic domain showing the N- and C-terminal lobes, an active site inhibitor, and the two Aspartes in the catalytic site.

BACE1 is mainly present in neurons, although it is expressed at low levels in most cell types of the body. This enzyme is expected to operate optimally at pH 4.0 -5.5 in acidic intracellular compartments such as the trans-Golgi network, endosomes and lysosomes, where it colocalizes with APP.^{35,36} BACE1 consists of three different structural domains: the N-terminal catalytic domain, the transmembrane domain and the cytosolic domain. The transmembrane and cytosolic domain anchor the protease to the membrane and regulate its cellular trafficking, respectively. The transmembrane domain also restricts BACE spatial distribution and

constrains the placement of the active site in the same orientation as APP.

The active site is more open and less hydrophobic than those of other human aspartic proteases, accommodating 12 pockets (P₄'-P₈).³⁷ Available crystal structures of inhibitors binding to BACE1 have demonstrated the inherent flexibility of this enzyme.³⁷ The catalytic center is covered by a flexible antiparallel β -hairpin loop between Val67 and Glu77 (called the 'flap') which is believed to control substrate access and orientation in the catalytic site.³⁸ The flexibility of the flap is the result of the different orientations of Tyr71, permitting two major conformations: the open or the closed conformation (Figure). The conformation of the protein is defined as "closed" when hydroxyl group of Tyr71 forms a hydrogen bond with the indole nitrogen of Trp76. On the other hand, the "open" conformation is the result of the movement of the tyrosine, which loses the hydrogen bond with Trp76 and forms a new hydrogen bond with the backbone carbonyl of Lys107. Thus moving the flap away from the catalytic center and allowing access of the substrates into the active-site cleft.³⁹ The flap region in the apo form of BACE1 is very flexible and frequently interchanges between the open and closed conformations.³⁷ However, upon ligand binding,

the flap region adopts the closed conformation, but the extent of closure depends on the ligand.

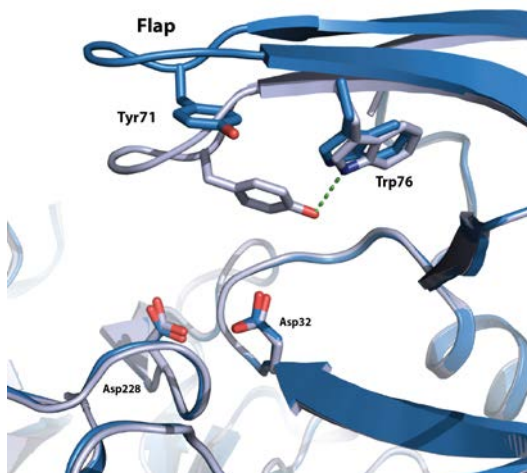


Figure 3. Closed (light blue; PDB 1W51) and open (dark blue; PDB 2OHU) conformations of BACE1 flap. Characteristic hydrogen bond between Tyr71 and Trp76 in the closed conformation is shown in green.

The BACE1 catalytic site contains another mobile structural feature termed by Patel *et al* as 10s loop (see Figure 4 panel A).⁴⁰ This short loop is located near the N-terminus between two β strands at the base of the S_3 pocket (residues 9-14) and displays three main low-energy conformations: a closed, an open and an outlier conformation, perfectly represented by PDB codes 1FKN, 1W51 and 1TQF, respectively (Figure panel B). Thus, the ligand

binding is modulated by the capability of the 10s loop to affect the shape of the S₃ pocket.

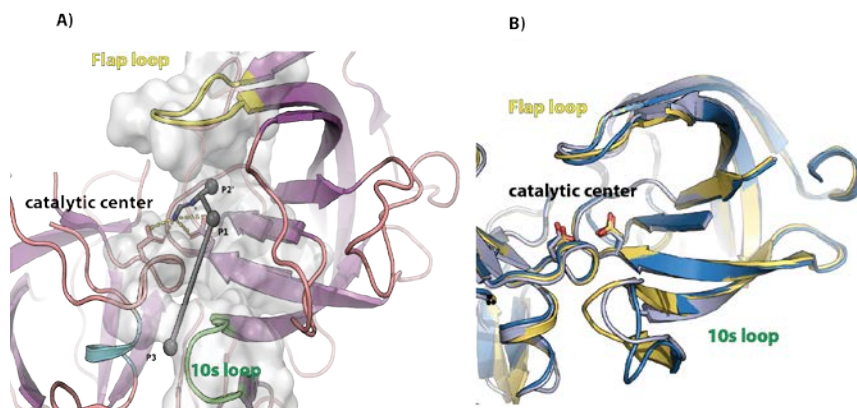


Figure 4. A) Schematic representation of the BACE1 pockets (P1, P2' and P3), the two flexible loops (Flap and 10s) and the catalytic center showing both aspartate residues. B) Closed (yellow), open (dark blue) and outlier (light blue) conformations of BACE1 10s loop.

4.5. BACE1 Inhibitors

Huge efforts are being made in academia and the pharmaceutical industry toward the design of BACE1 inhibitors, evidenced by the 370 BACE1 crystal structures available in the PDB,⁴¹. However, BACE1 has proven to be an exceptionally challenging target and pharmaceutical companies have not yet managed to commercialize any BACE1 inhibitor after more than 15 years of research.

First generation BACE1 inhibitors had peptidic nature and were initially designed as substrate analogs that mimicked the APP-cleavage sequence with a non-cleavable peptide bond, such as norstatine, statine, hydroxyethylene and hydroxyethylamine isosteres (Figure 5). They were un-druglike molecules.⁴² Subsequently attention turned to hydroxyethylene (HE) isosteres with fewer peptidic bonds. Replacement of the statine subunit with the HE motif resulted in an increased BACE1 inhibitory activity and improved cell permeability.

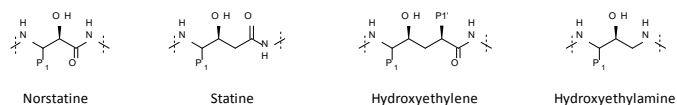


Figure 5. Transition state isosteres comprising a secondary hydroxyl group as peptide bond replacement.

In 2000 and 2001, Ghosh and Tang reported two HE derivatives: including the well-known OM99-2^{37,43} as a potent BACE1 inhibitor with $IC_{50} = 1.6$ nM (Figure 6). More simplified HE derivatives were ultimately found, but shifting to hydroxyethylamine (HEA) derivatives led to further improvements in potency and brain penetration, examples showing nanomolar activity in cellular assays

and reduction of amyloid in brain (Figure 7).^{44,45} Nevertheless these molecules were blighted with pharmacokinetic issues and P-glycoprotein (P-gp) efflux.

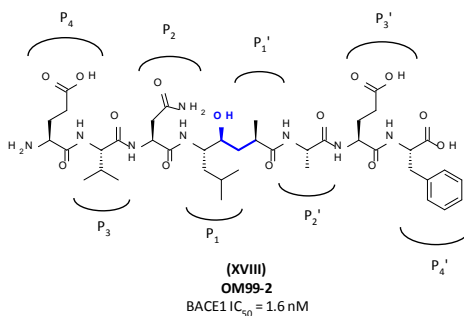


Figure 6. BACE1 inhibitor OM99-2.

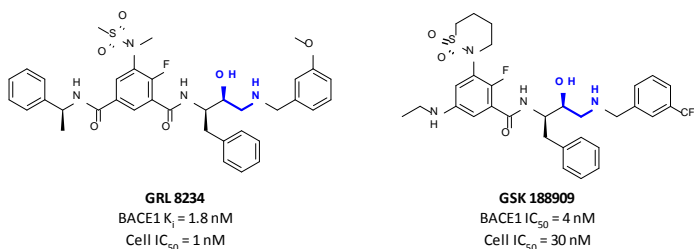


Figure 7. HEA inhibitors GRL-8234 and GSK 188909.

A second generation of BACE1 inhibitors presented lower molecular weight and polar surface area, better binding efficiency and drug-like properties, thus lowering P-gp efflux and increasing

central penetration. These series included carbinamine derivatives, macrocyclic inhibitors and amidine- or guanidine- class analogs. The carbinamine derivatives interact with the catalytic center of BACE1 through a single primary amine replacing the characteristic hydroxyl group from the HE framework, an example from Merck is shown in Figure 8, these two were strong P₃-gp substrates.⁴⁶ A macrocycle example, also Figure 8, had good potency (IC₅₀ = 2 nM) but poor brain penetration due to low permeability and high P-gp efflux, as well as poor pharmacokinetics.^{47,48} More recently, amidine and guanidine motifs have revolutionized the field of BACE1 inhibitors since they form a nice hydrogen-bonding network with the catalytic aspartyl dyad of BACE1.

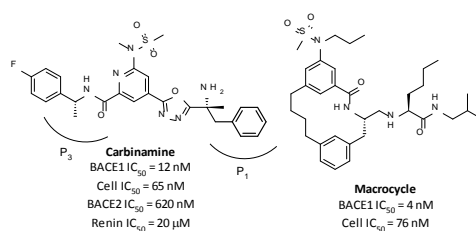


Figure 8. Example second generation BACE1 inhibitors.

High Throughput Screening and fragment based drug discovery campaigns identified the first amidine containing subclasses of

BACE1 inhibitors (Figure). Since the initial reports by Wyeth and Schering-Plough⁴⁹ in 2005, many pharmaceutical companies have elaborated and expanded on these amidine containing warheads trying to overcome the obstacles towards selective, safe and centrally efficacious BACE1 inhibitors. Years of research unveiled the relevance of monitoring the amidine pK_a when designing potent cell penetrant BACE1 inhibitors.^{50,51} There are many literature reports of medicinal chemistry optimization of these lead series, often discussing the issues of maintaining potency in P1, P3 and P2' pockets, whilst also optimizing PK considerations. Crystallographic X-ray structures are typically provided such as the example shown in Figure 10.⁵² Examples of amidine/guanidine BACE1 inhibitors have now entered clinical trials. LY-2811376 was the first amidine-like class BACE1 inhibitor to reach phase I, LY-2886721 was the second tested in the clinic, however it was discontinued due to abnormal liver function, and MK-8931 the most advanced with phase III started in November 2013, Figure 11.

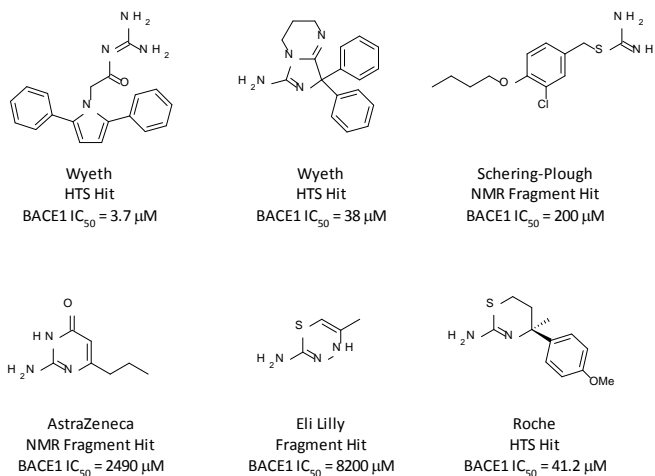


Figure 9. Representative amidine- and guanidine-like containing warheads identified by HTS and fragment screenings.



Figure 10. Crystal structures of BACE1 c complexed with two guanidine containing aminoimidazole BACE1 inhibitors.

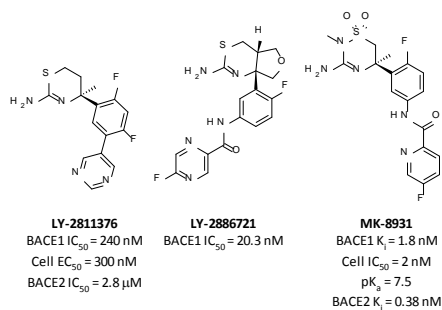


Figure 11. Clinical candidates from Eli-Lilly and Merck from the amidine/guanidine class of BACE1 inhibitors.

At present, whilst several amidine/guanidine class BACE1 inhibitors have entered clinical trials, some challenges remain, particularly with regards to selectivity versus enzymes such as BACE2. Also, with the pharmaceutical industry focusing its attention to amidine-like warheads the field is densely explored and very competitive, making the identification of novel binding motifs extremely challenging.⁵³ Therefore, it is still of interest to identify novel scaffolds or better understand issues of selectivity. In addition, the sheer volume of data and understanding a round BACE1 make it an ideal test case for methodologies that can later be applied in other drug discovery programs.

4.6. References

- 1 Organization, W. H. Alzheimer's Disease International. World Alzheimer Report. (2009).
- 2 Lahiri, D. K., Farlow, M. R., Greig, N. H. & Sambamurti, K. Current drug targets for Alzheimer's disease treatment. *Drug Development Research* **56**, 267-281, (2002).
- 3 Weisgraber, K. H., Rall, S. C. & Mahley, R. W. Human E apoprotein heterogeneity. Cysteine-arginine interchanges in the amino acid sequence of the apo-E isoforms. *Journal of Biological Chemistry* **256**, 9077-9083 (1981).
- 4 Amaducci, L. & Tesco, G. Aging as a major risk for degenerative diseases of the central nervous system. *Curr Opin Neurol* **7**, 283-286 (1994).
- 5 Fratiglioni, L., Ahlborn, A., Viitanen, M. & Winblad, B. Risk factors for late-onset Alzheimer's disease: A population-based, case-control study. *Annals of Neurology* **33**, 258-266, (1993).
- 6 Plassman, B. L. *et al.* Documented head injury in early adulthood and risk of Alzheimer's disease and other dementias. *Neurology* **55**, 1158-1166 (2000).
- 7 Sivanandam, T. M. & Thakur, M. K. Traumatic brain injury: A risk factor for Alzheimer's disease. *Neuroscience & Biobehavioral Reviews* **36**, 1376-1381, (2012).
- 8 Pendlebury, S. T. & Rothwell, P. M. Prevalence, incidence, and factors associated with pre-stroke and post-stroke dementia: a systematic review and meta-analysis. *The Lancet Neurology* **8**, 1006-1018, (2009).
- 9 Scarmeas, N. *et al.* Mediterranean Diet and Mild Cognitive Impairment. *Archives of neurology* **66**, 216-225, (2009).
- 10 Selkoe, D. J. Alzheimer's Disease: Genes, Proteins, and Therapy. *Physiological Reviews* **81**, 741 (2001).
- 11 Scheuner, D. *et al.* Secreted amyloid [beta]-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med* **2**, 864-870 (1996).
- 12 Goedert, M. & Spillantini, M. G. A Century of Alzheimer's Disease. *Science* **314**, 777 (2006).

- 13 Bartus, R. T., Dean, R. L., Beer, B. & Lippa, A. S. The cholinergic hypothesis of geriatric memory dysfunction. *Science* **217**, 408 (1982).
- 14 Perry, E. K. THE CHOLINERGIC HYPOTHESIS—TEN YEARS ON. *British Medical Bulletin* **42**, 63-69 (1986).
- 15 Francis, P. T., Palmer, A. M., Snape, M. & Wilcock, G. K. The cholinergic hypothesis of Alzheimer's disease: a review of progress. *Journal of Neurology, Neurosurgery & Psychiatry* **66**, 137-147 (1999).
- 16 Dumas, J. A. & Newhouse, P. A. The Cholinergic Hypothesis of Cognitive Aging Revisited A gain: Cholinergic Functional Compensation. *Pharmacology, biochemistry, and behavior* **99**, 254-261, (2011).
- 17 Cacabelos, R., Takeda, M. & Winblad, B. The glutamatergic system and neurodegeneration in dementia: preventive strategies in Alzheimer's disease. *International Journal of Geriatric Psychiatry* **14**, 3-47, (1999).
- 18 Parsons, C. G., Söffler, A. & Danysz, W. Memantine: a NMDA receptor antagonist that improves memory by restoration of homeostasis in the glutamatergic system - too little activation is bad, too much is even worse. *Neuropharmacology* **53**, 699-723, (2007).
- 19 Selkoe, D. J. The molecular pathology of Alzheimer's disease. *Neuron* **6**, 487-498, (1991).
- 20 Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184 (1992).
- 21 Roberts, G. W. *et al.* Beta amyloid protein deposition in the brain after severe head injury: implications for the pathogenesis of Alzheimer's disease. *Journal of Neurology, Neurosurgery, and Psychiatry* **57**, 419-425 (1994).
- 22 Kang, J. *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733-736 (1987).
- 23 Querfurth, H. W. & LaFerla, F. M. Alzheimer's Disease. *New England Journal of Medicine* **362**, 329-344, (2010).
- 24 Sisodia, S. S. & St George-Hyslop, P. H. [gamma]-Secretase, notch, A[beta] and Alzheimer's disease: Where do the presenilins fit in? *Nat Rev Neurosci* **3**, 281-290 (2002).
- 25 Sathya, M. *et al.* BACE1 in Alzheimer's disease. *Clinica Chimica Acta* **414**, 171-178, (2012).

- 26 Stockley, J. H. & O'Neill, C. Understanding BACE1: essential protease for amyloid- β production in Alzheimer's disease. *Cellular and Molecular Life Sciences* **65**, 3265, (2008).
- 27 Furukawa, K. *et al.* Increased Activity-Regulating and Neuroprotective Efficacy of α -Secretase-Derived Secreted Amyloid Precursor Protein Conferred by a C-Terminal Heparin-Binding Domain. *Journal of Neurochemistry* **67**, 1882-1896, (1996).
- 28 Morishima-Kawashima, M. & Ihara, Y. Alzheimer's disease: β -Amyloid protein and tau. *Journal of Neuroscience Research* **70**, 392-401, (2002).
- 29 Vassar, R. *et al.* β -Secretase Cleavage of Alzheimer's Amyloid Precursor Protein by the Transmembrane Aspartic Protease BACE. *Science* **286**, 735 (1999).
- 30 Yan, R. *et al.* Membrane-anchored aspartyl protease with Alzheimer's disease [beta]-secretase activity. *Nature* **402**, 533-537 (1999).
- 31 Sinha, S. *et al.* Purification and cloning of amyloid precursor protein [beta]-secretase from human brain. *Nature* **402**, 537-540 (1999).
- 32 Hussain, I. *et al.* Identification of a Novel Aspartic Protease (Asp 2) as β -Secretase. *Molecular and Cellular Neuroscience* **14**, 419-427, (1999).
- 33 Lin, X. *et al.* Human aspartic protease memapsin 2 cleaves the β -secretase site of β -amyloid precursor protein. *Proceedings of the National Academy of Sciences* **97**, 1456-1460 (2000).
- 34 Creemers, J. W. M. *et al.* Processing of β -Secretase by Furin and Other Members of the Proprotein Convertase Family. *Journal of Biological Chemistry* **276**, 4211-4217 (2001).
- 35 Capell, A. *et al.* Maturation and Pro-peptide Cleavage of β -Secretase. *Journal of Biological Chemistry* **275**, 30849 - 30854 (2000).
- 36 Yan, R. & Vassar, R. Targeting the β secretase BACE1 for Alzheimer's disease therapy. *The Lancet Neurology* **13**, 319-329, (2014).
- 37 Hong, L. *et al.* Structure of the Protease Domain of Memapsin 2 (β -Secretase) Complexed with Inhibitor. *Science* **290**, 150 (2000).

- 38 Shimizu, H. *et al.* Crystal Structure of an Active Form of BACE1, an Enzyme Responsible for Amyloid β Protein Production. *Molecular and Cellular Biology* **28**, 3663-3671, (2008).
- 39 Spronk, S. A. & Carlson, H. A. The role of tyrosine 71 in modulating the flap conformations of BACE1. *Proteins* **79**, 2247-2259, (2011).
- 40 Patel, S., Vuillard, L., Cleasby, A., Murray, C. W. & Yon, J. Apo and Inhibitor Complex Structures of BACE (β -secretase). *Journal of Molecular Biology* **343**, 407-416, doi:http://dx.doi.org/10.1016/j.jmb.2004.08.018 (2004).
- 41 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
- 42 Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **23**, 3-25, (1997).
- 43 Ghosh, A. K. *et al.* Structure-Based Design: Potent Inhibitors of Human Brain Memapsin 2 (β -Secretase). *Journal of Medicinal Chemistry* **44**, 2865-2868, (2001).
- 44 Ghosh, A. K. *et al.* Potent memapsin 2 (β -secretase) inhibitors: Design, synthesis, protein-ligand X-ray structure, and in vivo evaluation. *Bioorganic & Medicinal Chemistry Letters* **18**, 1031-1036, (2008).
- 45 Hussain, I. *et al.* Oral administration of a potent and selective non-peptidic BACE-1 inhibitor decreases β -cleavage of amyloid precursor protein and amyloid- β production in vivo. *Journal of Neurochemistry* **100**, 802-809, (2007).
- 46 Rajapakse, H. A. *et al.* Discovery of Oxadiazoyl Tertiary Carbinamine Inhibitors of β -Secretase (BACE-1). *Journal of Medicinal Chemistry* **49**, 7270-7273, (2006).
- 47 Lindsley, S. R. *et al.* Design, synthesis, and SAR of macrocyclic tertiary carbinamine BACE-1 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **17**, 4057-4061, (2007).
- 48 Stachel, S. J. *et al.* Macrocyclic Inhibitors of β -Secretase: Functional Activity in a Animal Model. *Journal of Medicinal Chemistry* **49**, 6147-6150, (2006).

- 49 Wang, Y.-S. *et al.* Application of Fragment-Based NMR Screening, X-ray Crystallography, Structure-Based Design, and Focused Chemical Library Design to Identify Novel μM Leads for the Development of nM BACE-1 (β -Site APP Cleaving Enzyme 1) Inhibitors. *Journal of Medicinal Chemistry* **53**, 942-950, (2010).
- 50 Tresadern, G. *et al.* Rational design and synthesis of aminopiperazinones as β -secretase (BACE) inhibitors. *Bioorganic & Medicinal Chemistry Letters* **21**, 7255-7260, (2011).
- 51 Rombouts, F. J. R. *et al.* 1,4-Oxazine β -Secretase 1 (BACE1) Inhibitors: From Hit Generation to Orally Bioavailable Brain Penetrant Leads. *Journal of Medicinal Chemistry* **58**, 8216-8235, (2015).
- 52 Malamas, M. S. *et al.* Aminoimidazoles as Potent and Selective Human β -Secretase (BACE1) Inhibitors. *Journal of Medicinal Chemistry* **52**, 6314-6323, (2009).
- 53 Oehlich, D., Prokopova, H. & Gijzen, H. J. M. The evolution of a midine-based brain penetrant BACE1 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **24**, 2033-2045, (2014).

4.7. The Application of Free Energy Perturbation for the Design of BACE1 Inhibitors

This Chapter is a collaboration with Myriam Ciordia, Francisca Delgado, Andrés A. Trabanco and Gary Tresadern from Janssen Pharmaceutical Research. This work is currently published at *J Chem Inf Model* 2016, 56, 1856-1871.

4.7.1 Background

Alzheimer's disease (AD) is a neurodegenerative illness that chronically affects multiple brain functions and causes changes in the behavior of the sufferer, often leading to death within 3 to 9 years after diagnosis. Despite the high prevalence of AD, and it being the most common form of dementia (contributing to 60-70% of cases),¹ there is no cure and currently available drugs only provide a modest delay on the decline of cognitive function.² Hence, there are considerable research efforts to intervene in disease progression.³⁻⁵ Inhibition of β -secretase 1 (BACE1) is the most compelling approach. The hypothesis is based on slowing or preventing the cleavage of amyloid precursor protein (APP) into the neurotoxic $A\beta_{42}$ peptide products that make up the core of the

amyloid plaques seen in the AD brains.⁶ Furthermore, genetic evidence has reinforced the attention of academic and industrial researchers on BACE1 for the development of AD therapeutics.^{7,8}

First generation BACE1 inhibitors had peptidic character and were initially designed as substrate analogues that mimicked the APP-cleavage sequence with a non-cleavable peptide bond. Despite high *in vitro* potency they often had undesirable physicochemical properties for a CNS drug and typically showed poor oral bioavailability and difficulty to cross the blood brain barrier (BBB).⁹⁻¹¹ The discovery of the amidine moiety revolutionized the field of BACE1 inhibitors by allowing access to a second generation of non-peptidic derivatives that form an ideal hydrogen-bonding network with the catalytic aspartyl dyad (Asp32 and Asp228) of the enzyme, Figure 1.^{12,13} Whilst the work of others has shown the importance of this motif,¹⁴ within our labs we first identified this interaction via a benzoguanidine series¹⁵ and have subsequently explored various alternative chemical series.¹⁶⁻¹⁹ This generation of BACE1 inhibitors often contains a quaternary sp³ carbon that provides an ideal vector for the substituents to fill the P1-P3 and P2' pockets of the catalytic site, Figure 1.^{20,21}

A major goal of computational chemistry is the accurate prediction of protein–ligand binding affinities.²² Free-energy simulations provide a rigorous approach and methods such as free-energy perturbation (FEP), thermodynamic integration (TI), and λ dynamics, use molecular dynamics or Monte Carlo simulations to compute the free-energy difference between two structurally related ligands.²³ During drug discovery lead optimization (LO) it is normally required to explore the chemical space around key lead molecules via the synthesis of close analogues. Hence, computation of accurate relative binding affinities (i.e., the difference in binding energy between two analogue compounds) is of high interest and can make a significant impact in drug design, whilst also avoiding the computationally challenging prediction of absolute binding free energies. The calculation of protein–ligand binding affinities in this manner dates back over thirty years.²⁴⁻³⁰ More recently, improved force fields, new sampling algorithms, and low-cost parallel computing (often graphics processing units GPU), have improved accuracy and turnaround time needed to impact LO efforts in various academic projects.³¹⁻³³ However, reports of the impact in industrial pharmaceutical research programs are scarce but beginning to emerge.^{34,35} The significant therapeutic interest in

BACE1 has prompted many computational studies. For instance, work has ranged from the catalytic mechanism and the optimal protonation state of the catalytic aspartates,³⁶⁻³⁸ the flexibility of the active site and value of ensemble docking,³⁹ towards more recently the accurate estimation of binding energies *via* quantum mechanics (QM).⁴⁰

In this work, efforts to optimize a novel series of amidine containing spirocyclic BACE1 inhibitors are described. In particular, the latest generation of FEP implementation is used to predict binding energies in a retrospective and prospective manner. We continue our exploration of a spiroaminodihydropyrrole scaffold. We recently reported a series of C₃-containing aminopyrrolidines as BACE1 inhibitors with moderate to good activity. Initial SAR studies identified **1** as the most potent, Figure 2.¹⁸ Interestingly, modification of the synthetic route for each aminopyrrolidine inhibitors like **1** allowed access to a highly novel spirocyclic core that was used as a template for the design of new examples represented by derivative **2**. There are only a few reports describing similar spirocyclic scaffolds such as **3** and **4** (Figure 2) likely due to their challenging synthesis.⁴¹⁻⁴³ However, AZD3293 (**4**)^{44,45} is currently in phase II and is the most advanced BACE1

inhibitor containing a spirocyclic warhead, suggesting these types of scaffolds offer promise.¹ Surprisingly, our initial spirocyclic prototypes such as **2** were inactive. Hence, here we explore alternative decoration using FEP to design and prioritize molecules with substituents to fill the P1-P3 pockets. We first report a retrospective comparison of FEP predicted binding energies and experiment for a similar spirocyclic series. Then we performed prospective FEP binding energy predictions on a set of 18 molecules from our spiroaminodihydropyrrole scaffold and a subset were synthesized based on their results. Our results show good correlation between predicted and experimental binding energies, providing further evidence that FEP can be used as a tool to assist lead optimization, even for BACE1 which is considered a difficult and structurally flexible target.⁴⁶ The FEP approach outperformed docking and MM-GBSA methods. This study therefore provides a valuable contribution describing FEP as a tool for drug design.

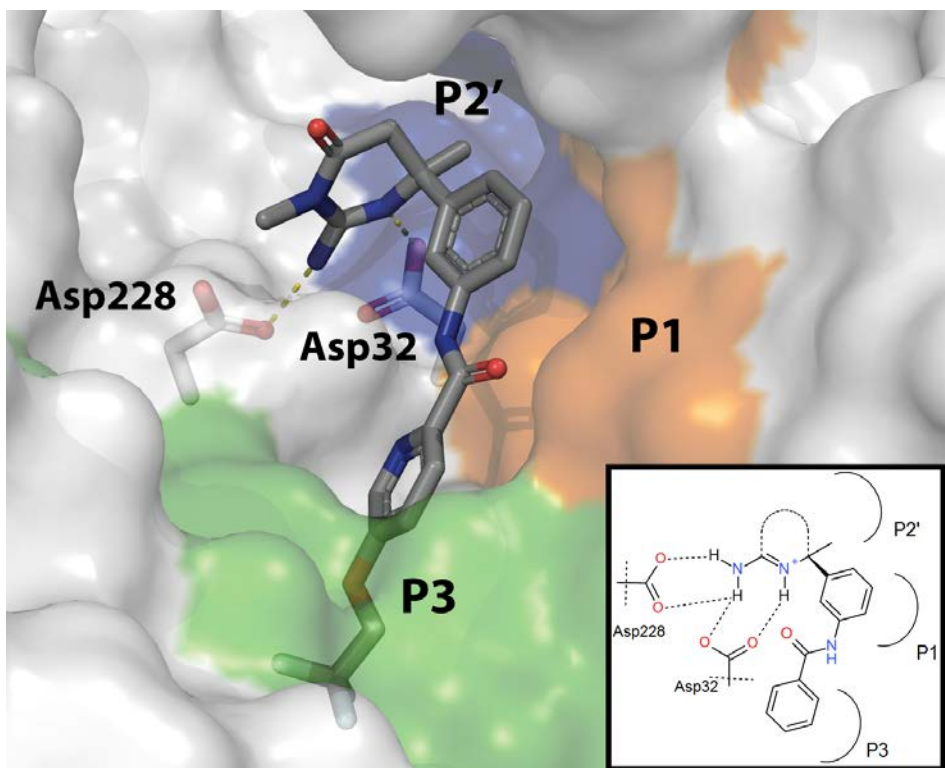


Figure 1. Schematic binding mode in the BACE1 active site showing the key H-bond and salt bridge interaction between the amidine and the catalytic aspartates (also see inset), and sub-pockets are highlighted with surface coloring P2' blue, P1 orange and P3 green. The figure was generated from PDB structure 3ZOV.⁴⁷

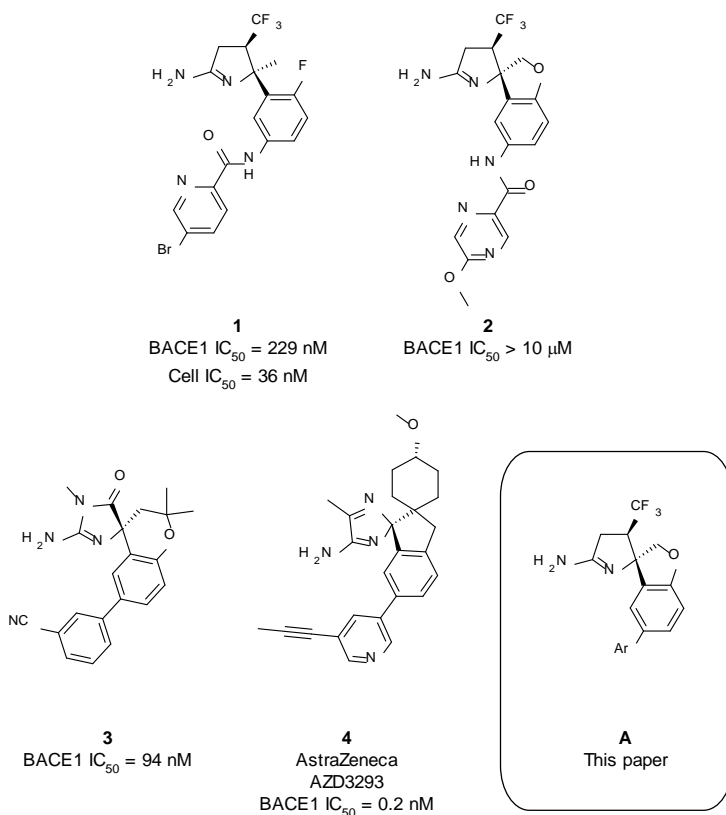


Figure 2. Examples of spirocyclic BACE1 inhibitors.

4.7.2. Results

The performance of FEP on spirocyclic BACE1 inhibitors was first tested in a retrospective manner using 32 molecules taken from the work of Hunt *et al*⁴³, Table 1. A crystal structure of the 3-cyanophenyl R-group (PDB 4JPC) was used to place the molecules in the binding site, see supporting information for input binding modes. The recommended FEP+ approach is a single 5 ns

simulation for each of the twelve λ windows, hence 60 ns cumulative simulation time per perturbation. Here FEP+ refers to the Schrödinger Inc. FEP implementation based on OPLSv3, DESMOND GPU MD, REST sampling and cycle closure error corrections, see experimental section for details. The default mapper identified 62 chemical perturbations between analogues based on considerations such as structural similarity, see supporting information Figure S 1. We also performed 1, 10 and 20 ns simulations, as such, 12, 120 and 240 ns cumulative simulation time per perturbation. All FEP perturbations were performed in both solvent and protein.

The resulting predicted dG (kcal/mol) are compared to experiment in Table 1 and Figure 3. In brief, the results for all simulation times can be considered good (below an error of 1 kcal/mol)³⁴ with MUE compared to experiment of 0.71 ± 0.18 , 0.58 ± 0.15 , 0.57 ± 0.12 and 0.57 ± 0.11 kcal/mol for 1, 5, 10 and 20 ns simulations respectively. The correlation with experiment is almost identical in each case (Figure 3). Also, the correlation between FEP simulation times was high, for instance the R^2 between binding energy predictions from 5 and 20 ns simulations was 0.95. The cycle closure error is a calculated parameter that helps to understand

the reliability of the predicted binding energies. It determines to what extent the sum of the calculated free energies deviate from zero for each closed thermodynamic cycle within the FEP+ mapper. The maximum and mean cycle closure error for the 5 ns simulations were 0.71 and 0.34 kcal/mol respectively. Whereas for the 20 ns simulations the values were 0.38 and 0.22 kcal/mol suggesting improved internal consistency due to longer simulation times. Looking in more detail at Table 1 reveals interesting results for specific compounds. The 5-(prop-1-yn-1-yl)pyridin-3-yl substituent for instance showed a big improvement in MUE with increasing simulation time: 3.03, 2.42, 1.83 and 1.21 kcal/mol at 1, 5, 10 and 20 ns respectively. This was also seen for the 3-ethoxyphenyl substituent that had an MUE of 1.71, 0.89, 0.46 and 0.02 kcal/mol at the four corresponding simulation times. This was in contrast to the majority of molecules that displayed fluctuations in MUE but with no apparent trend. Hence, the molecules that were the biggest outliers versus experiment at 1 ns benefited from increased simulation time. This dataset of 32 molecules was also submitted to docking and MM-GBSA calculations delivering worse performance compared with the FEP calculations. In this case Glide SP and XP docking delivered R^2 correlation coefficients with experiment of

0.48 and 0.37. Two MM-GBSA protocols, one without a active site minimization and one using minimization of an 8 Å radius around the ligands, provided R^2 correlation of 0.33 and 0.01 respectively. Overall, the retrospective application suggested FEP would be suitable for exploring the P1-P3 pocket substituents and also that some molecules benefit from increased simulation time.

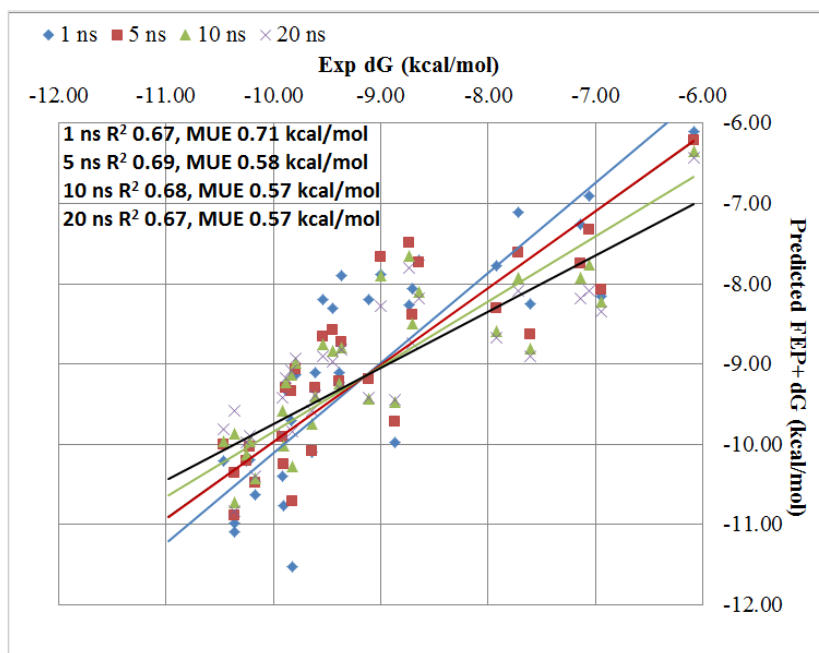


Figure 3. Correlation between FEP+ predicted and experimental binding energy for the retrospective application. Results from 1, 5, 10 and 20 ns simulation time FEP calculations are shown. See Tables 1 and 3 for more details.

We now discuss the prospective design of new spirocyclic scaffold BACE1 inhibitors. Using computational modelling we

determined that spirocyclic molecules with amide linkers between the phenyl substituent in P1 and the distal aromatic in P3 (such as **2**) were too long, resulting in clash between the ligand and enzyme.¹⁸ Compounds of general structure **A** (Figure 2), where the amide linker was deleted, would allow for the distal aromatic ring to occupy the P1-P3 pockets without clashing with the protein. This was tested by performing conformational analysis on **1**, **2**, and our prototype **A** with a distal phenyl included as the Ar group. The resultant accurate set of low energy conformers were placed in the BACE1 binding site in such a way to maintain optimal interaction with the catalytic aspartates, Figure 4. The spirocycle forces an ‘orthogonal’ or T-shaped orientation of the substituent on the sp³ carbon of the amidine heterocycle. Therefore long groups such as biaromatic amides clash with the binding site surface. This is not the case for active molecule **1** for which some conformations can adopt a more optimal shape, similar to the crystallographic ligand. Also, for the spirocyclic scaffold, a shorter biaryl substituent such as prototype **A** fits in the binding site whilst avoiding clash with the enzyme. The distal Ar group (Ph in this case) is well placed to permit *meta* substituents to enter deeper into the P3 pocket.

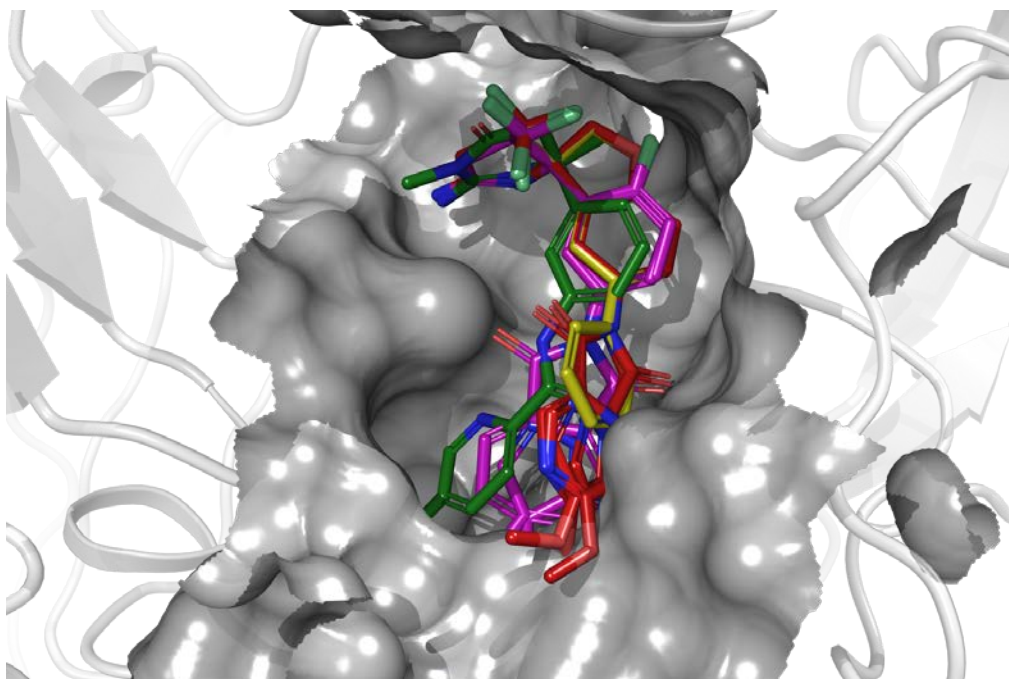


Figure 4. Conformational analysis of molecules **1** (purple), **2** (red) and prototype **A** (yellow) with distal phenyl substitution. The ligand from crystal structure 3ZOV (green, same molecule as Figure 1) is shown for comparison of an optimal fit in the binding site. Conformers are shown in the binding site of BACE1 to display the clash with the surface for molecules of type **2** versus the ability for **1** to adopt a binding pose similar to the crystal structure or **A** to be free from clash.

The eighteen molecules (Figure 5) were docked and the preferred binding pose was used as input for the FEP+ calculations. The input binding modes are provided in supporting information. The FEP+ mapper generated 26 underlying pairwise perturbations (Figure S2 and Table S1). An additional 12 connections were added manually for some pairs of molecules to ensure all compounds were involved

in at least 3 perturbations, hence the complete mapper included 38 perturbations. Adding additional connections can provide improved binding energy estimates for molecules with fewer connections on the perimeter of the mapper. It can also improve error estimates by the cycle closure method although in this case the extra perturbations had little effect.^{48,49} For the most part similarity scores were high, reflecting the suitability of this set of molecules for FEP+ calculations.⁵⁰

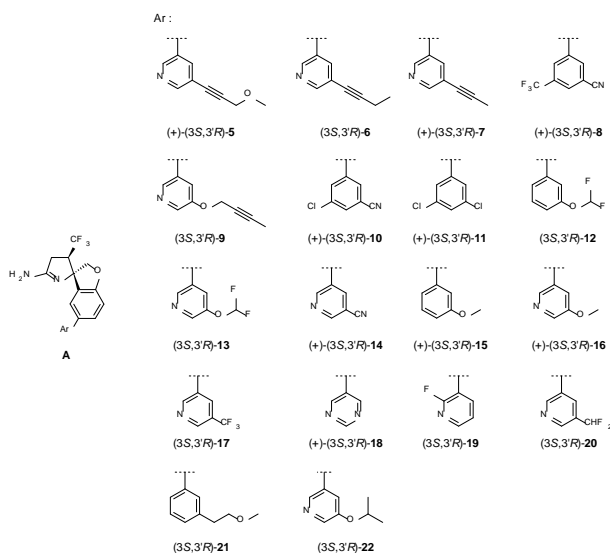


Figure 5. Set of 18 proposed Ar groups coupled to scaffold A.

An initial 5 ns simulation was performed and the results are shown in Figure 6 and Table 2. In general, the cycle closure errors for each predicted binding free energy difference were good,

ranging from 0.33 to 0.65 with a mean of 0.53 kcal/mol (Table 3). Molecules **5** and **6** were predicted to be the most active, having a calculated binding energy 4.66 and 4.01 kcal/mol more stable than **15**. Molecules such as **17** and **18** were predicted to be less favourable binders, with ΔG 0.26 and 0.59 kcal/mol higher in energy than **15**. Hence a broad range in predicted binding energies was seen across these compounds. The FEP+ predicted error (Bennett error) is a net error estimate arising from the underlying Bennett acceptance ratio (BAR) theory, see experimental section for more details. Thus, the FEP+ predicted error had a maximum of 0.89 with a mean of 0.65 kcal/mol (Table 3). The results from this simulation, based on the default recommended approach, were used to make the selection of molecules to synthesize.

Repeat 5 ns calculations were performed two more times using new random seeds to test the stability of the results. The results based on the average for the three simulations are shown in Tables 2 and 3. The correlation coefficient between the single 5 ns simulation and the average of three separate calculations was $R^2 = 0.97$, confirming very similar results. The standard errors in measurements were generally small, ranging from 0.15 to 0.31

kcal/mol. Overall, the averaged results from the repeat calculations did not diverge greatly from the single 5 ns simulation.

As in the retrospective application, calculations were performed for 1, 10 and 20 ns for each λ step of the perturbations for all molecules shown in Table 1. The results and performance statistics are summarized in Tables 2 and 3. The protein and ligand RMSD for the end point λ -replicas from the 20 ns simulations are provided in supporting information Table S1. They confirm that the inhibitor remained in its bound pose throughout the simulations with no large protein or ligand movements seen. We have not observed difficulties of this type for FEP B ACE1 calculations and the retrospective calculations were similarly stable in this respect (data not shown). For the 1 ns simulations the cycle closure errors for each predicted binding free energy difference ranged from 0.49 to 0.91 with a mean of 0.79 kcal/mol. Whereas for the 10 ns simulation times the cycle closure errors for each perturbation ranged from 0.40 to 0.77 with a mean of 0.56 kcal/mol. In turn, the cycle closure errors for the 20 ns simulations ranged from 0.26 to 0.53 with a mean of 0.45 kcal/mol. As seen for the retrospective application, these results show improvement in errors with increasing simulation time. Likewise this is seen for the predicted

Bennett error associated with each molecule's estimated dG. The R² correlation coefficient between the single 5 ns final predicted FEP+ dG and the 1, 10 and 20 ns results was 0.96, 0.95 and 0.94 respectively. Hence a gain as seen in the retrospective application, we observed good stability of the predicted dG with respect to simulation time. This suggests that the extended simulation time had little impact on the overall outcome of the calculations. However, the predicted errors improved with increasing simulation time. This is an important factor because lower errors provide additional confidence that is important to initiate chemical synthesis.

Based on the results of the single FEP+ calculation run at 5 ns MD simulation time (default approach) a selection of 9 molecules were synthesized, encompassing examples predicted to be most active, such as **5** and **7** but also molecules predicted to be less potent, such as **16** and **18**. We chose molecules across the range of predicted activity and with alternative Ar group substitution but also that would be synthetically feasible given the difficulties to reach spirocyclic BACE1 inhibitors of this type. The synthetic route developed to access spiroaminodihydropyrroles (*3S,3'R*)-**5-18** is depicted in Scheme 1. Firstly, Michael addition-cyclization

methodology as previously described afforded intermediate (\pm)-**(2*S*,3*R*)-24**. Then, subsequent nitration, reduction and intramolecular nucleophilic aromatic substitution led to (\pm)-**(3*S*,3'*R*)-27**. This intermediate was then reduced to the corresponding aniline (\pm)-**(3*S*,3'*R*)-28** followed by replacement of the amino group by a bromine to yield (\pm)-**(3*S*,3'*R*)-29**. The amidine (\pm)-**(3*S*,3'*R*)-31** was reached in a two-step sequence involving thionation and treatment with aqueous ammonia. Finally, the nine boronic acids selected from the FEP predictions (Table 2) were coupled via Suzuki reaction, delivering the final products (\pm)-**(3*S*,3'*R*)-5-18**.

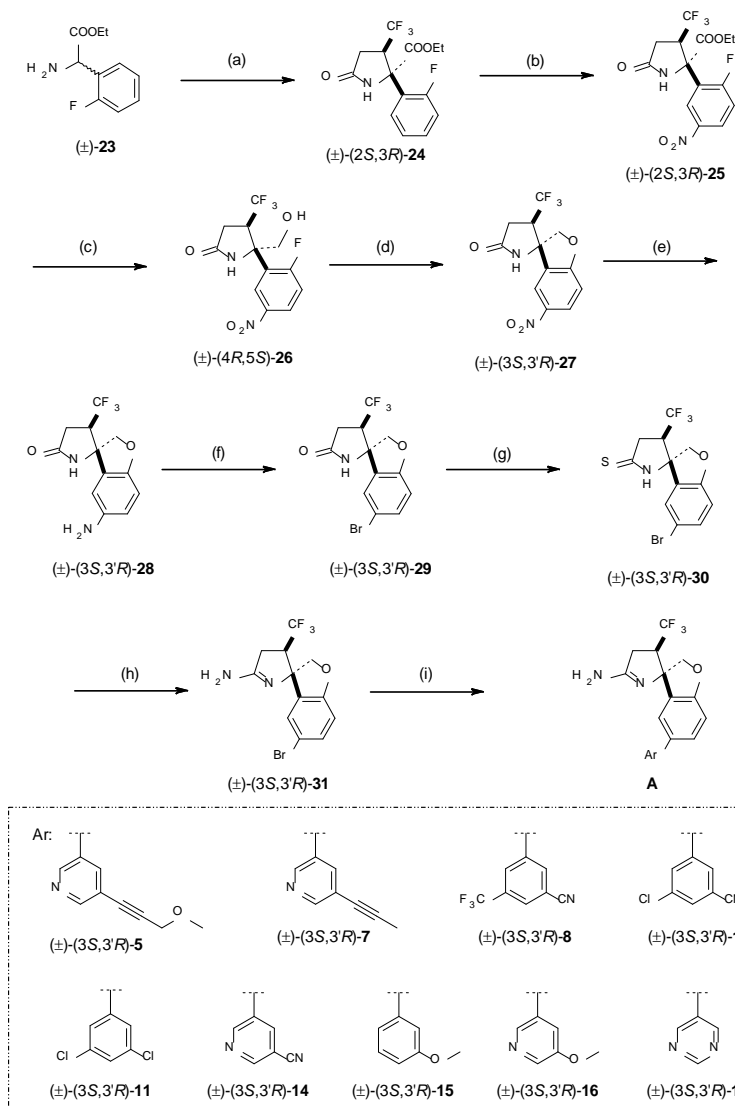
Compounds were screened in a BACE1 inhibition assay and the resulting pIC₅₀ data are provided in Table 2. Chiral SFC separation of the racemates allowed the isolation of both enantiomers. Only the amidines with absolute configuration (*3*S*,3'*R**) (alloft hem dextrorotatory) were active in the enzymatic assay, hence (+) nomenclature used here.⁵¹ The final molecules that were tested versus BACE1 and reported in Table 2 were all in the (+)-**(3*S*,3'*R*)** form, although this full naming is omitted in much of the discussion below. The molecules were considerably more potent than our previous spiroaminodihydropyrrole containing an amide linker (**2**),

but they can only be considered moderately active compared to other optimized BACE1 inhibitors such as **3** and **4**. The most potent example (+)-(3*S*,3'*R*)-**7** had a BACE1 inhibition pIC₅₀ of 6.94.

The overall correlation between predicted and experimental ΔG (kcal/mol) is shown in Figure 6 with details provided in Tables 2 and 3. For a relatively small number of molecules, or for a small range of binding energies, the R^2 can be less reliable and instead the errors (MUE) in prediction can be of more use.³⁴ Correlation for the dataset run with a single 5 ns simulation time was R^2 0.45 with a mean unsigned error of 0.91 ± 0.49 kcal/mol. One large outlier affects the results. Molecule **5** was predicted to be the most active of all examples, but turned out to have only a moderate potency, with pIC₅₀ 6.19. Interestingly, the results calculated as the average of three separate 5 ns simulations, showed better R^2 and MUE compared to experiment, with values of 0.54 and 0.86 ± 0.41 kcal/mol respectively. Considering modification of the simulation time, the R^2 and MUE for the 1 ns simulation time were 0.52 and 1.04 ± 0.48 kcal/mol respectively, whereas, with 10 ns simulation they were 0.64 and 0.71 ± 0.34 kcal/mol. For the 20 ns simulation the corresponding R^2 and MUE were 0.68 and 0.59 ± 0.29 kcal/mol. Docking methods performed worse than the longest FE

simulations. Glide SP and XP docking showed R^2 correlation with experiment of 0.11 and 0.35 respectively. This is as expected for methods not designed to predict relative binding affinities of congeners and consistent with recent reports.⁵² Also, a simple cLogP metric exhibited an R^2 of 0.31, see supporting information Figure S 3. MM-GBSA calculations based on the FEP starting geometries and using a default approach without protein flexibility provided an R^2 of 0.08. Whereas, MM-GBSA using a flexible region around the binding site delivered an R^2 of 0.35, see Figure S4 in supporting information. These comparative results were similar to a recent study that used the same FEP+ implementation to predict relative binding energies of fragment-like molecules.⁵³ Hence, FEP, particularly with longer simulation times, performed better than conventional methods and averaging over repeats and using longer simulation time ultimately delivered lower errors compared with experiment.

Scheme 1. Synthesis of spiroaminodihydropyrroles^a



^a**Reagents and conditions:** (a) ethyl 4,4,4-trifluoro-trans-2-butenate, NaH, THF, rt, 6 h, 62 %; (b) HNO₃, H₂SO₄, 0 °C, 30 min, 92 %; (c) NaBH₄, THF, H₂O, rt, 1 h, 78 %; (d) NaH, DMF, rt, 1 h, 53 %; (e) H₂, Pd/C, MeOH, rt, 16 h, 57 %; (f) isoamyl nitrite, CuBr₂, CH₃CN, 65 °C, 5 h, 70 %; (g) P₂S₅, THF, 80 °C, 16 h, 96 %; (h) NH₃ aq. and NH₃ 7 N in MeOH, 110 °C, 1 h, microwave, 91 %; (i) ArB(OH)₂, Pd(PPh₃)₄, NaHCO₃, 1,4-dioxane, 70 °C, 16 h, 26-71 %.

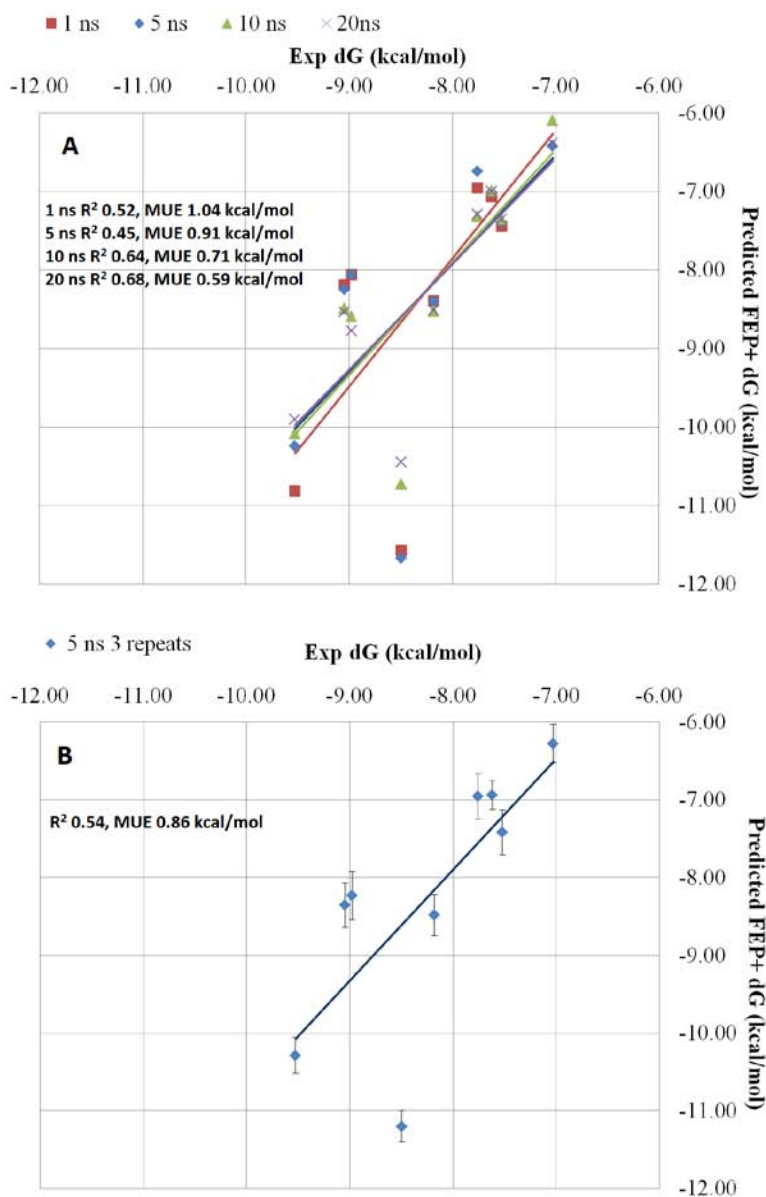


Figure 6. Correlation between FEP+ predicted and experimental binding energy for the prospective application. A) Results from single 1, 5, 10 and 20 ns simulation time FEP+ calculations and experimental dG (kcal/mol) for the synthesized set of BACE inhibitors. B) Results for three repeats of 5 ns simulation, error bars show the standard error in measurement for calculated binding energy, see Tables 2 and 3.

4.7.3. Discussion and conclusions

The predicted binding energies from 1, 10 and 20 ns simulations were highly correlated to the 5 ns results. However, the cycle closure error reduced with increasing simulation time for both applications. Prior to synthesis we cannot know the performance with respect to experiment, but observing improved stability in the cycle closure error with increasing simulation time provides confidence that errors associated with insufficient sampling are likely to be eliminated. The results showed good correlation for the predicted binding energies and experiment. In the prospective application, using repeat 5 ns simulations improved the correlation and error versus experiment when compared with a single 5 ns simulation. Increasing the simulation time also delivered better R^2 and MUE statistics for the prospective application but did little to improve the results for the retrospective dataset beyond the 5 ns simulation time. The FEP method outperformed docking and MM-GBSA approaches in both retrospective and prospective applications. However, some caution is needed because these results were not significant within 90 % confidence intervals. The sampling

provided by the FEP method appears most beneficial for a subset of the molecules in both datasets, as discussed below.

Looking in more detail at the MUE it is apparent that there is an improvement in the predicted binding energy for outliers such as the 3-ethoxyphenyl and 5-(prop-1-yn-1-yl)pyridin-3-yl in Table 1 and 5 in Table 2 when increasing the simulation time. In addition, the predicted binding energies for **10** and **11** were also improved for the 20 ns simulations compared to 1 and 5 ns. The majority of compounds such as **8**, **14**, and **15**, and many in the retrospective application, showed no improvement in predicted vs experimental binding when increasing the simulation time. These compounds contained smaller less flexible P3 substituents and were converged at 5 ns therefore not benefiting from additional sampling. Hence, the improvement due to increased sampling time was not uniform for all molecules and overall the results at different simulation times remained highly correlated. However, the benefit of extra sampling for some molecules was seen in the cycle closure error which reduced with longer simulations, confirming the value of this parameter to understand the reliability of the predictions ahead of synthesis. Improvement due to sampling for a subset of molecules has been seen before in free energy calculations on drug discovery

datasets.⁵⁴ The improvement for only a minority of molecules explains why statistically significant differences in the MUE between simulation times were not seen. For example, considering a 90% confidence interval in the prospective application, the 1 ns FEP results showed MUE of 1.01 ± 0.48 kcal/mol, whereas the 20 ns FEP simulations had MUE of 0.59 ± 0.29 kcal/mol, Table 3.

In the prospective application, the pyridines bearing the larger acetylenic substituents provided higher predicted activity, such as **7** that was experimentally the most active. The binding mode for these compounds features an H-bond between the pyridyl nitrogen to a conserved water molecule that bridges to Gly11 (numbered as Gly59 in PDB 4JPC structure),⁴³ Figure 7. This conformationally restrains the distal aromatic ring, and permits substituents *meta* to the pyridyl nitrogen to access a deeper P3 subpocket going towards Ser229 (Ser277 in PDB 4JPC) and displace water molecules in the process (compare to crystal structure PDB 4JPE⁴³ for instance). Hence the orientation of this heteroaromatic ring is restrained on one side by the H-bond and on the other by the substituent accessing the P3 pocket (Figure 7A). Overall, this provides a relatively high degree of confidence in placing the molecules at the start of the FEP+ calculations in what is likely to be their correct

binding mode. This was also the case for most molecules in the retrospective application that only featured small structural changes compared to the crystal structure. Molecule **5** has a similar acetylenic substituent *meta* to the pyridyl nitrogen, but in this case it contains a larger and more flexible distal methoxymethylene group. This molecule was the largest outlier from both FEP+ applications. It was predicted to be the most potent compound, approximately 0.9 kcal/mol more potent than **7**, but instead it was less potent by approximately 1 kcal/mol. The quality of the prediction benefited from increased simulation time (Figure 6A, Table 2). However, the standard error in measurement for this compound in the 5 ns repeat simulations (Figure 6B) was not significantly different from the other molecules in this study, suggesting it did not behave differently during new repeats. Given that the binding energy prediction for this compound improved with increased sampling it is plausible that the initial starting geometry for the molecule or protein was suboptimal (Figure 7B). The flexible distal methoxymethylene was in close proximity to the protein surface and may require the whole Ar ring to rotate to an alternative dihedral orientation that may be challenging to sample given the size of the substituent. Also, this substituent approaches the 10s loop of the P3

pocket. This is known to be a region of conformational flexibility,⁴⁶ hence, the protein may also need to adapt to this particular group and may do so on a slow timescale that benefits from extra sampling. Hence, the implication of this work is that where uncertainty about exact binding mode exists, longer simulation times than 5 ns should be considered.

Regarding the practicalities of using FEP+ for molecular design, one main consideration is to maintain pace with chemistry and provide recommendations in a timely way. In this study, the calculations for a 5 ns perturbation in the protein complex took ~12 h running on 4 NVIDIA Tesla K20m GPUs and ~3.5 h in solvent. In turn, 1, 10, and 20 ns perturbations in the protein complex took ~3, ~24 and ~46 h respectively. We studied 32 and 18 compounds and performed a total of 62 and 38 perturbations for our two applications and multiple FEP+ runs. Running the default 5 ns binding energy predictions for the prospective application on a single 4 GPU K20 node would take approximately 589 h. At the time of performing this work we had access to 4 such nodes and results could be attained in approximately 147 h (~6 days). This was an acceptable time frame for this study, meaning results could be discussed amongst the medicinal chemistry team within a week.

The longer simulation times and larger numbers of compounds clearly imply increased computational cost. In this work we performed approximately 48 microseconds of combined MD simulations. When applying FEP+ in other projects we are often calculating larger sets of virtual molecules, up to 60, 70 or 80 for instance. This offers considerable advantages to identify stand out molecules with exceptionally high predicted activity, as well as allowing many closer analogues to be calculated, hence providing more reliable closer structural perturbations. As such we continue to invest in GPU hardware to permit running larger sets of molecules with FEP+.

The recent report of Wang *et al*³⁴ described multiple FEP+ applications in drug discovery projects. One of the examples therein was an application on BACE1. They studied 36 molecules (58 perturbations) and reported a MUE of 0.84 kcal/mol that is in agreement with our work for a single 5 ns simulation (MUE 0.91 kcal/mol). The authors also reported an average MUE for applications on eight different targets of approximately 0.9 kcal/mol. Another recent study by scientists from Pfizer applied FEP in a prospective manner to evaluate 17 potential Spleen Tyrosine Kinase (Syk) Inhibitors. Although only two examples

from the calculations were synthesized the results were in agreement with one being predicted more and the other less potent than the reference compound.⁵⁵ Along with other well-known studies, such as design of non-nucleoside inhibitors of HIV-1 reverse transcriptase,⁵⁶ these recent applications and our work here show the potential of FEP for drug discovery and lead optimization. Given the current high interest to attain the “Holy Grail” of accurate binding energy predictions within 1 kcal/mol of experiment⁵⁷ we expect continued research and activity in this area.

Overall we have demonstrated a retrospective and prospective application of FEP+ for the design of new BACE1 inhibitors. The retrospective application studied 32 molecules and showed good correlation between predicted and experimental binding energies, which was largely consistent with respect to simulation time.

The longest simulation time provided R^2 correlation and MUE compared to experiment of 0.67 and 0.57 ± 0.11 kcal/mol. A small number of outliers benefited from extra sampling. The prospective calculations were performed on a set of 18 molecules and subsequently 9 examples were synthesized. Good agreement with calculated and experimental activities was also seen, that improved with increasing simulation time, also due to the reduction in errors

for difficult outliers until reaching R^2 correlation and MUE compared to experiment of 0.68 and 0.59 ± 0.29 kcal/mol for the longest simulation performed on all molecules, in the same range as the retrospective application on a structurally similar dataset. Prior to synthesis the cycle closure errors can be studied with respect to simulation time thereby providing increased confidence for compound prioritization. The trend in MUE with increasing simulation time was not statistically significant for the whole dataset. Such results are well within acceptable accuracy for use in prospective molecular design, arguably preferred to be better than 1 kcal/mol.^{53,57} Small improvements were also seen when using multiple repeat calculations with the same simulation time and averaging the results. The FEP calculations, particularly with longer simulation times, provided better results than alternative docking and MM-GBSA approaches for both retrospective and prospective datasets. This work provides further evidence of the value of FEP for molecular design, particularly when having confidence in the underlying binding mode and sufficient computational sampling.

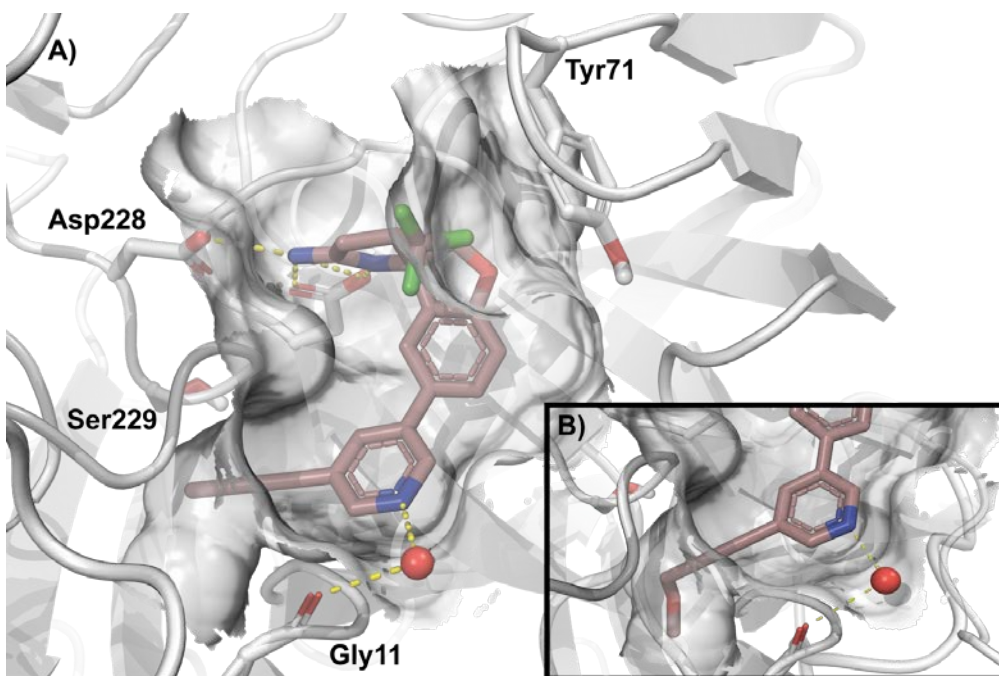
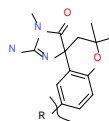
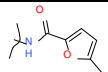


Figure 7. Binding mode of specified compounds. A) **7** showing orientation of distal aromatic ring, water bridged H-bond interaction to Gly11 and other selected amino acids. B) **5** showing a close up of the P3 pocket and the close fit of the methoxymethylene acetylenic group against the protein surface.

Table 1. The retrospective application of FEP+ to predict the binding energies of a dataset of existing spirocyclic BACE1 inhibitors taken from Hunt *et al* Calculations were performed for 1, 5, 10 and 20 ns simulation time.

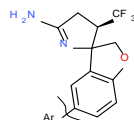


R	Calculations: FEP+ dG (kcal/mol) ^a				Experimental ^b		MUE (kcal/mol) ^d			
	1 ns	5 ns	10 ns	20 ns	BACE1 pIC ₅₀	dG (kcal/mol) ^c	1 ns	5 ns	10 ns	20 ns
Br	-6.11	-6.21	-6.35	-6.44	4.43	-6.09	0.02	0.12	0.26	0.35
Phenyl	-8.06	-8.38	-8.50	-8.74	6.34	-8.71	0.64	0.32	0.21	0.03
3-chlorophenyl	-10.21	-10.00	-9.97	-9.81	7.62	-10.46	0.25	0.46	0.49	0.65
3-cyanophenyl	-10.10	-10.08	-9.75	-9.58	7.03	-9.65	0.45	0.44	0.10	0.07
3-methoxyphenyl	-9.70	-9.33	-9.14	-9.08	7.17	-9.84	0.14	0.51	0.70	0.76
2-methoxyphenyl	-7.77	-8.30	-8.59	-8.67	5.77	-7.92	0.15	0.38	0.67	0.75
4-methoxyphenyl	-6.91	-7.32	-7.76	-8.09	5.14	-7.06	0.15	0.26	0.70	1.03
3,5-dichlorophenyl	-11.09	-10.88	-10.73	-10.84	7.55	-10.37	0.72	0.51	0.36	0.47
3-fluoro-5-methoxyphenyl	-10.76	-10.25	-10.02	-9.88	7.21	-9.90	0.86	0.35	0.11	0.03
2-fluoro-5-methoxyphenyl	-10.40	-9.90	-9.58	-9.42	7.22	-9.91	0.49	0.01	0.33	0.50
3-chloro-5-fluorophenyl	-10.63	-10.48	-10.43	-10.40	7.41	-10.17	0.46	0.31	0.26	0.23
5-chloro-2-fluorophenyl	-10.19	-10.03	-9.96	-9.90	7.44	-10.22	0.03	0.19	0.26	0.32
3-isopropoxyphenyl	-8.25	-8.63	-8.81	-8.91	5.54	-7.61	0.64	1.02	1.20	1.30
3-ethoxyphenyl	-11.53	-10.71	-10.28	-9.85	7.15	-9.82	1.71	0.89	0.46	0.02
3-(trifluoromethyl)phenyl	-9.13	-9.07	-8.99	-8.93	7.13	-9.79	0.66	0.72	0.80	0.86
3-(methylthio)phenyl	-9.98	-9.71	-9.47	-9.45	6.46	-8.87	1.11	0.84	0.60	0.57
3-(difluoromethoxy)phenyl	-10.98	-10.36	-9.87	-9.59	7.55	-10.37	0.61	0.01	0.50	0.78
3-fluorophenyl ^e	-9.11	-9.30	-9.41	-9.39	7.00	-9.61	0.50	0.31	0.21	0.23
Cyclohexyl	-7.88	-7.66	-7.90	-8.28	6.56	-9.00	1.12	1.34	1.10	0.72
-OCH ₂ CH(Me) ₂	-8.26	-7.49	-7.66	-7.80	6.37	-8.74	0.48	1.25	1.08	0.94
Piperidin-1-yl	-8.16	-8.07	-8.23	-8.35	5.06	-6.95	1.21	1.12	1.28	1.40
Pyridin-3-yl	-8.19	-8.66	-8.75	-8.90	6.95	-9.54	1.35	0.88	0.79	0.64
Pyridin-2-yl	-7.11	-7.61	-7.92	-8.09	5.63	-7.72	0.61	0.11	0.19	0.36

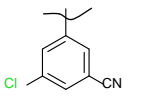
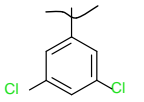
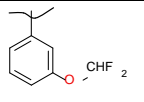
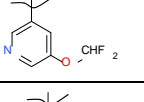
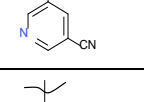
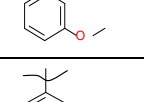
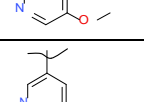
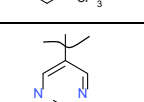
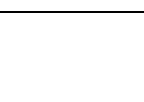
Pyridin-4-yl	-7.26	-7.74	-7.92	-8.19	5.20	-7.14	0.12	0.60	0.78	1.04
5-chloropyridin-3-yl	-10.21	-10.21	-10.13	-9.98	7.47	-10.25	0.04	0.04	0.12	0.28
5-(trifluoromethyl)pyridin-3-yl	-9.32	-9.29	-9.23	-9.18	7.20	-9.88	0.56	0.59	0.66	0.71
5-(prop-1-yn-1-yl)pyridin-3-yl	-14.01	-13.40	-12.81	-12.20	8.00	-10.98	3.03	2.42	1.83	1.21
2-fluoropyridin-3-yl	-8.30	-8.58	-8.83	-8.98	6.89	-9.45	1.15	0.87	0.62	0.48
5-fluoropyridin-3-yl	-9.10	-9.21	-9.25	-9.30	6.84	-9.39	0.29	0.18	0.14	0.09
Pyrimidin-5-yl	-7.90	-8.73	-8.80	-8.83	6.82	-9.36	1.46	0.63	0.56	0.54
Isothiazol-5-yl	-7.71	-7.73	-8.10	-8.18	6.30	-8.65	0.94	0.92	0.55	0.47
	-8.19	-9.18	-9.44	-9.42	6.64	-9.11	0.92	0.07	0.33	0.30

Footnotes: ^a FEP estimated error from the calculation in kcal/mol in parenthesis. ^b Taken from the reference of Hunt *et al* based on conversion of differences in pIC₅₀ into kcal/mol, T = 300K. ^d based on normalising the FEP+ relative dG to the experimental dG by mean centered difference (kcal/mol), see Table 3 for summary of average MUE for each of the FEP+ calculations. ^e 3-fluorophenyl was the reference for the FEP calculations.

Table 2. The prospective application of FEP+ to predict the binding energies of newly designed BACE1 inhibitors. Calculations were performed for 1, 5, 10 and 20 ns simulation time on the set of 18 molecules, including results for the average of three repeats of 5 ns simulations.



Cp d.	Ar	Calculations: FEP+ dG (kcal/mol) ^a					Chosen for synthesis	Experimental		MUE (kcal/mol) ^e				
		1 ns	5 ns ^b	5 ns ^c	10 ns	20 ns		BACE1 pIC ₅₀	dG (kcal/mol) ^d	1 ns	5 ns ^b	5 ns ^c	10 ns	20 ns
5		-11.57 (1.03)	-11.66 (0.75)	-11.20 (0.97) [0.20]	-10.72 (0.72)	-10.44 (0.59)	Yes	6.19	-8.50	3.07	3.16	2.70	2.22	1.94
6		-11.46 (1.16)	-11.01 (0.73)	-10.96 (0.99) [0.15]	-10.86 (0.84)	-10.53 (0.62)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
7		-10.81 (0.98)	-10.23 (0.8)	-10.29 (0.81) [0.23]	-10.08 (0.65)	-9.90 (0.60)	Yes	6.94	-9.53	1.28	0.70	0.76	0.55	0.37
8		-8.39 (1.06)	-8.39 (0.73)	-8.48 (0.84) [0.26]	-8.52 (0.74)	-8.51 (0.72)	Yes	5.96	-8.18	0.21	0.21	0.30	0.34	0.33
9		-9.19 (1.24)	-8.32 (0.89)	-8.41 (1.10) [0.26]	-8.57 (0.82)	-8.28 (0.68)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.

10		-8.19 (1.13)	-8.24 (0.7)	-8.36 (0.82) [0.29]	-8.48 (0.88)	-8.53 (0.71)	Yes	6.59	-9.05	0.86	0.81	0.69	0.57	0.52
11		-8.06 (1.00)	-8.05 (0.72)	-8.23 (0.72) [0.31]	-8.59 (0.75)	-8.77 (0.65)	Yes	6.54	-8.98	0.92	0.93	0.75	0.39	0.21
12		-8.24 (0.90)	-7.98 (0.63)	-7.79 (0.80) [0.15]	-7.71 (0.65)	-7.73 (0.53)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
13		-8.05 (0.49)	-7.77 (0.56)	-7.55 (0.67) [0.20]	-7.40 (0.40)	-7.47 (0.34)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
14		-7.44 (0.87)	-7.39 (0.51)	-7.42 (0.80) [0.29]	-7.35 (0.77)	-7.35 (0.53)	Yes	5.48	-7.52	0.08	0.13	0.10	0.17	0.17
15		-7.07 (0.00)	-7.00 (0.00)	-6.94 (0.19) [0.19]	-6.99 (0.0)	-7.00 (0.0)	Yes	5.55	-7.62	0.55	0.62	0.68	0.63	0.62
16		-6.67 (1.20)	-6.74 (0.66)	-6.78 (0.94) [0.21]	-7.15 (0.77)	-7.22 (0.68)	Yes	5.65	-7.76	1.09	1.02	0.98	0.61	0.54
17		-6.95 (1.40)	-6.74 (0.83)	-6.95 (0.91) [0.29]	-7.31 (0.91)	-7.29 (0.77)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
18		-5.69 (0.87)	-6.41 (0.63)	-6.28 (0.45) [0.25]	-6.09 (0.53)	-6.38 (0.51)	Yes	5.12	-7.03	1.34	0.62	0.75	0.94	0.65

19		-5.61 (1.16)	-6.40 (0.65)	-6.19 (0.82) [0.25]	-6.29 (0.90)	-6.43 (0.63)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
20		-6.25 (1.16)	-6.33 (0.74)	-6.43 (0.69) [0.30]	-6.92 (0.75)	-6.81 (0.69)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
21		-6.49 (0.73)	-6.10 (0.55)	-5.95 (0.74) [0.16]	-5.91 (0.57)	-5.92 (0.42)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.
22		-5.90 (0.81)	-6.07 (0.56)	-5.94 (0.67) [0.21]	-6.12 (0.53)	-6.32 (0.50)	No	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.	n.m.

Footnotes: All molecules had (+)-(3*S*,3'*R*) chirality. n.m. not measured. Molecules named with '+' showed positive optical rotation, whereas those which were not synthesized could not be analysed. All (3*S*,3'*R*) enantiomers were used for the computational work. ^a FEP estimated error from the calculation in kcal/mol in parenthesis. Compound **15** was the reference for the FEP calculations. ^b FEP results from the first single 5 ns simulation. ^c FEP results based on average from three separate 5 ns simulations, standard error in measurement provided in square parentheses. ^d based on conversion of differences in pIC₅₀ into kcal/mol, T = 300K. ^e based on normalising the FEP+ relative dG to the experimental dG by mean centered difference (kcal/mol), see Table 3 for summary of average MUE for each of the FEP+ calculations.

Table 3. Summary of performance statistics for the different FEP+ calculations.

FEP+ calculations:	Cycle Closure Error (kcal/mol)			FEP+ predicted error (kcal/mol)		Comparison with experiment	
	Min	Max	Mean	Max	Mean	R ² correlation coefficient	MUE ^a (kcal/mol)
Retrospective application on known BACE1 inhibitors (32 molecules, 62 perturbations)							
1 ns	0.08	0.59	0.36	0.80	0.62	0.67	0.71 ± 0.18
5 ns	0.06	0.71	0.34	0.98	0.66	0.69	0.58 ± 0.15
10 ns	0.06	0.53	0.33	0.83	0.62	0.68	0.57 ± 0.12
20 ns	0.05	0.38	0.22	0.65	0.50	0.67	0.57 ± 0.11
Prospective application for design of new BACE1 inhibitors (18 molecules of which 9 synthesized, 38 perturbations)							
1 ns	0.49	0.91	0.79	1.40	0.96	0.52	1.04 ± 0.48
5 ns	0.33	0.65	0.53	0.89	0.65	0.45	0.91 ± 0.49
5 ns 3 repeat averaged	0.39	0.88	0.69	1.10	0.77	0.54	0.86 ± 0.41
10 ns	0.40	0.77	0.56	0.91	0.68	0.64	0.71 ± 0.34
20 ns	0.26	0.53	0.45	0.77	0.57	0.68	0.59 ± 0.29

^a The mean unsigned error is provided with 90 % confidence intervals

4.7.4. Experimental section

Computational conformational analysis. Conformational search was performed with Mixed torsional/Low mode sampling available in Maestro v2015-3, using the OPLSv3 force field,⁵⁸ solvent and other parameters set at default. Conformers within 10 kcal/mol from the global minimum were passed to QM minimization with Jaguar at the B3LYP level of theory using the LACVP** (necessary because of the Bromine atom in **1**) that uses a 6-31G** basis with an effective core potential for the Br, solvation was included with the Poisson Boltzmann Finite (PBF) water model. To exclude very similar conformations, QM minimized structures within 0.2 Å RMSD of another were removed.

Computational docking procedure. Docking was performed using the Glide software (Release 2015-3) from Schrödinger. All BACE1 Protein Data Bank⁵⁹ (PDB) structures were searched to identify a structure with ID 4JPC⁴³ that had a spirocyclic ligand structurally similar to those under study herein. The protein was prepared for docking as follows. Firstly, 4JPC was imported into Maestro⁶⁰ and structure preparation was performed using the Protein Preparation Wizard⁶¹ with default settings to fix missing

sidechains/atoms, assign protein protonation states with PROPKA,⁶² optimize the hydrogen bonding network, assign ligand charges, and relax crystal contacts with a brief minimization to RMSD 0.5 Å. The catalytic aspartates were treated in their unionized states. The ligand crystalized in 4JPC was used to place the docking grid box. All active site waters were retained and treated as part of the receptor grid. Two H-bond constraints on the Asp from the catalytic center (Asp289 and Asp93, often referred to as Asp228 and Asp32, respectively) were chosen to perform the docking. The eighteen ligand molecules were prepared for docking using the LigPrep tool. All default settings were used except ionization was explicitly set to ensure all ligands were protonated on the amino pyrrolidine ring. The ligands were parameterized for use with the OPLS force field up front using the tools available in Maestro, for partial charges and newly calculated force field parameters see supporting information. ConfGen⁶³ with fast settings was used to derive multiple 3D conformers for each molecule. All conformers were then passed as input to the Glide XP docking thereby producing multiple docking solutions for each conformer of each molecule. The Glide XP scoring function was used, but sampling was increased through modifying a number of parameters within Glide: expanded

sampling was turned on, and 15 initial poses were passed to post-docking minimization. All other docking parameters were set to the defaults. Results were then aggregated at the level of each molecule and the best poses inspected. For performance comparisons Glide SP docking and MM-GBSA calculations were also performed. Glide SP docking was run with default settings, a new grid of 20 Å centered on Asp80 (4JPC amino acid numbering), and permitting ligands up to 20 Angstroms in length. The MM-GBSA calculations were run with the same XP docking poses used as input for the subsequent FEP calculations. The VSGB solvation model was used along with force field minimization of the ligand and a default approach with no protein minimization and a second approach with an 8 Å radius of the surrounding binding site (using the same active region for all ligands).

Computational FEP procedure. All calculations were conducted using version 2015-3 of the Schrödinger molecular modeling suite. The FEP methodology used here combines an accurate modern force field, OPLSv3 (with parameterisation for each ligand calculated upfront),⁶⁴ the efficient GPU-enabled parallel molecular dynamics engine Desmond version 3.9,⁶⁵ the REST enhanced sampling technique^{66,67} and the cycle-closure

correction algorithm⁶⁸ to incorporate redundant information into free energy estimates, it is often referred to as FEP+. Calculations were conducted using the FEP+ mapper technology⁴⁸ to automate setup and analysis. Desmond uses soft core potentials to overcome possible van der Waals endpoint instabilities at λ 0 and 1. Overall default computation protocols were used with a 5 ns simulation length for ligands both in complex and in solution. The 5 ns simulation on the prospective dataset was repeated to compare with data based on the average of three separate calculations. Each repeat used a different random seed to provide alternative random velocities to start each MD simulation. In addition we performed 1, 10 and 20 ns simulations for comparison with the recommended 5 ns default approach. These additional and longer simulations were performed for all molecules, not just outliers or any smaller subset. Also of note, these additional simulations were run independently from the beginning (0 ns), not as extensions of the 5 ns simulations for instance. As mentioned above, the FEP+ calculations were performed with the molecules in the chosen docked pose in the 4JPC BACE1 crystal structure as starting conformation. Molecules were treated in an ionized form and missing force field parameters were calculated up front. Proteins were prepared as described above

for the docking calculations using the Protein Preparation Wizard in Maestro. All resolved crystal water molecules were retained for the free energy simulations. The results of the simulations with 5 ns simulation time were used to define the molecules recommended for synthesis in the prospective application. We report theoretical error estimates based on cycle closure methodology, the theoretical FEP+ predicted error and also the mean unsigned error compared to experiment. The cycle closure error assesses the reliability of the predictions by determining how much the sum of the calculated free energy changes, for each closed thermodynamic cycle within the FEP+ mapper, deviates from the theoretical value of 0.⁶⁸ The FEP+ theoretical error (Bennett error) is derived from the Bennett acceptance ratio (BAR) analytical error as the square root of the estimated variance of the total free energy.⁶⁸⁻⁷⁰

Enzymatic BACE1 assay. BACE1 enzymatic activity was assessed by a FRET assay using an amyloid precursor protein (APP) derived 13 amino acids peptide that contains the ‘Swedish’ Lys-Met/Asn-Leu mutation of the APP beta-secretase cleavage site as a substrate (Bachem cat No. M-2465) and soluble BACE1(1-454) (Aurigene, Custom made). This substrate contains two fluorophores, (7-methoxycoumarin-4-yl) acetic acid (Mca) is a

fluorescent donor with the excitation wavelength at 320 nm and emission at 405 nm and 2,4-dinitrophenyl (Dnp) is a proprietary quencher/acceptor. The increase in fluorescence is linearly related to the rate of proteolysis. In a 384-well format, BACE1 is incubated with the substrate and the inhibitor. The amount of proteolysis is directly measured by fluorescence measurement in the Fluoroskan microplate fluorometer (Thermo scientific). For the low control no enzyme was added to the reaction mixture.

Chemistry.

Synthetic Route: The synthetic route developed to access spiroaminodihydropyrroles (*3S,3'R*)-**5-18** is depicted in Scheme 1. Firstly, taking advantage of the one-pot diastereoselective tandem Michael addition-cyclization methodology for the synthesis of polyfluorosubstituted pyrrolidones we previously described,¹⁸ intermediate (\pm)-(*2S,3R*)-**24** was prepared by reaction between (\pm)-**23** and ethyl 4,4,4-trifluoro-trans-2-butenolate. Then, regioselective nitration in the *para* position to the aromatic fluorine using a mixture of HNO₃/H₂SO₄ afforded (\pm)-(*2S,3R*)-**25** in excellent yield. Subsequent reduction of the ester group with NaBH₄ provided alcohol (\pm)-(*4R,5S*)-**26** in almost quantitative yield. Due to the presence of the nitro group, the activated aromatic fluorine in (\pm)-

(4*R*,5*S*)-**26** easily underwent the intramolecular nucleophilic aromatic substitution by attack of the alcohol using NaH as base, generating (±)-(3*S*,3'*R*)-**27**. With intermediate (±)-(3*S*,3'*R*)-**27** in hand, reduction of the nitro group to the corresponding aniline (±)-(3*S*,3'*R*)-**28** by hydrogenation, followed by replacement of the amino group by a bromine via Sandmeyer reaction afforded (±)-(3*S*,3'*R*)-**29**. Synthesis of amidine (±)-(3*S*,3'*R*)-**31** from amide (±)-(3*S*,3'*R*)-**29** was performed by a two-step sequence involving thionation and treatment with aqueous ammonia under microwave irradiation. Nine boronic acids selected from the proposed sets (Figure 4) were coupled to (±)-(3*S*,3'*R*)-**31** via Suzuki reaction, allowing access to final products (±)-(3*S*,3'*R*)-**5-18**.

General Methods. Unless otherwise noted, all reagents and solvents were obtained from commercial suppliers and used without further purification. Thin layer chromatography (TLC) was carried out on silica gel 60 F₂₅₄ plates (Merck). Flash column chromatography was performed on silica gel, particle size 60 Å, mesh of 230–400 (Merck), under standard techniques. Microwave assisted reactions were performed in a single-mode reactor: Biotage Initiator Sixty microwave reactor (Biotage) or in a multimode reactor: MicroSYNTH Labstation (Milestone, Inc.). Nuclear

magnetic resonance (NMR) spectra were recorded with either a Bruker DPX-400 or a Bruker AV-500 spectrometer (Bruker AG) with standard pulse sequences operating at 400 and 500 MHz, respectively, using CDCl₃ and DMSO-*d*₆ as solvents. Chemical shifts (δ) are reported in parts per million (ppm) downfield from tetramethylsilane ($\delta = 0$). Coupling constants are reported in hertz. Splitting patterns are defined by s (singlet), d (doublet), dd (double doublet), t (triplet), q (quartet), or m (multiplet). High resolution mass spectra were recorded on a Q-ToF mass spectrometer configured with an electrospray ionization source, maintained at 140 °C, using nitrogen as the nebulizer gas, argon as collision gas and Lockmass device for mass calibration using Leucine-Enkephaline as standard substance. Spectra were acquired either in positive or in negative ionization mode, by scanning from 50 to 1200 Da in 0.1 s. In positive mode, the capillary needle voltage could vary from 0.25 to 2.0 kV. In negative mode, the capillary needle voltage was 2.0 kV. Cone voltage was 25 V in both ionization modes. Optical rotations were measured on a Perkin-Elmer 341 polarimeter with a sodium lamp and reported as follows: $[\alpha]_D^T(\lambda, c \text{ g}/100 \text{ mL}, \text{ solvent})$. Melting points (mp) were determined with a DSC823e (Mettler-Toledo) apparatus and measured with a

temperature gradient of 10 °C/min. Maximum temperature was 300 °C. Peak values were recorded. Values are peak values and are obtained with experimental uncertainties that are commonly associated with this analytical method. Purities of all new compounds were determined by analytical RPHPLC using the area percentage method on the UV trace recorded at a wavelength of 254 nm and were found to have $\geq 95\%$ purity unless otherwise specified.

Synthetic protocols. *General Synthetic Procedure for the preparation of compounds (\pm) -(3*S*,3'*R*)-5-18.* Pd(PPh₃)₄ (0.15 equiv) was added to a degassed suspension of (\pm) -(3*S*,3'*R*)-**31** (1.0 equiv), the corresponding arylboronic acid (1.5 equiv) and NaHCO₃ aq. sat. solution (0.69 M) in 1,4-dioxane (0.28 M) and the suspension was stirred at 70 °C for 16 h. The mixture was allowed to cool to rt and the solvent was evaporated *in vacuo*. The crude product was purified by flash column chromatography (dry load) (silica gel; 7 M solution of ammonia in MeOH in DCM, from 0/100 to 4/96). The desired fractions were collected and concentrated *in vacuo* to yield the corresponding analogue (\pm) -(3*S*,3'*R*)-**5-18**. The racemates (\pm) -(3*S*,3'*R*)-**5-18** were purified by chiral SFC to afford $(+)$ -(3*S*,3'*R*)-**5-18** and $(-)$ -(3*R*,3'*S*)-**5-18**.

4.7.5. Supporting information

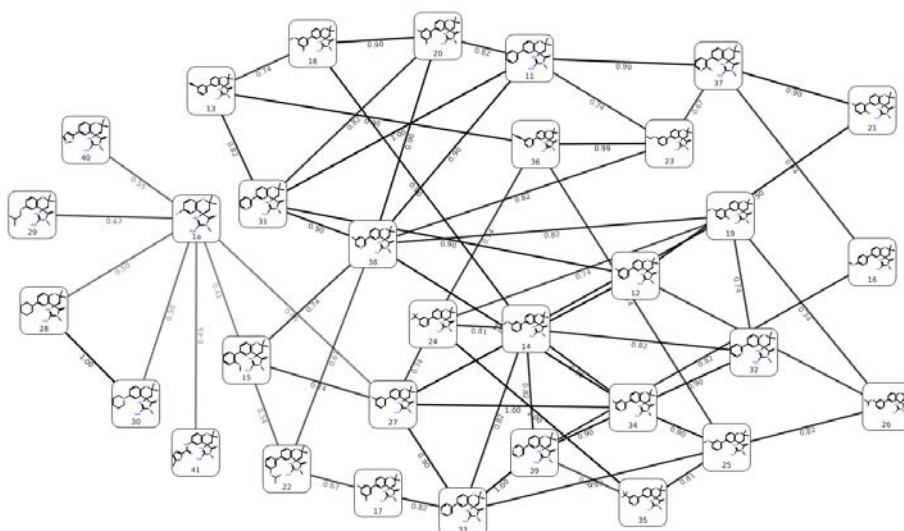


Figure S1. The FEP+ mapper output showing the molecular perturbations performed in the calculations for the retrospective application. The numbers on each edge represent the similarity score between the pairs of molecules.

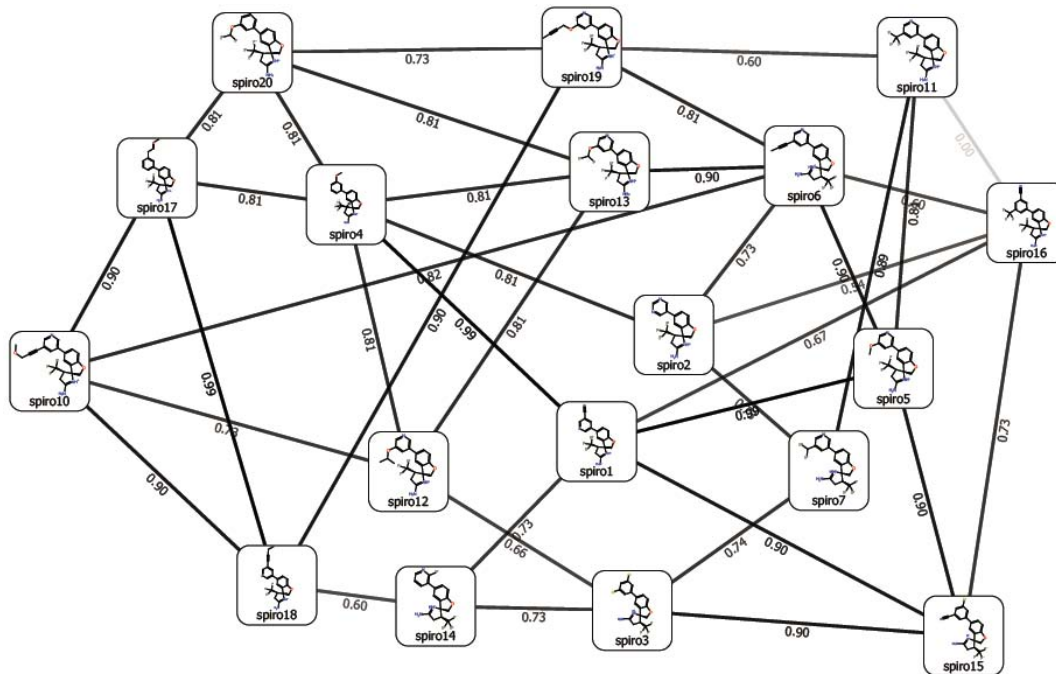
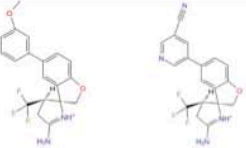
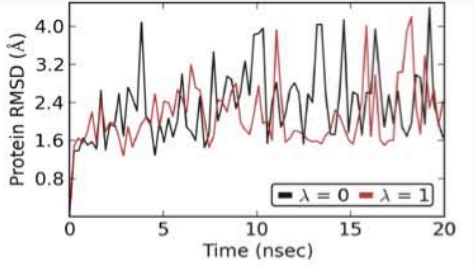
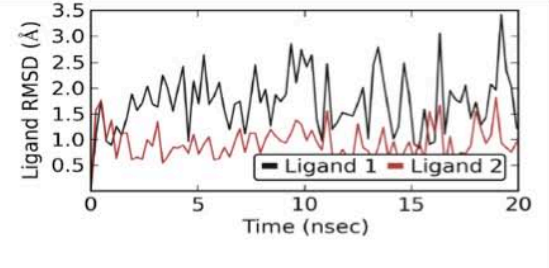
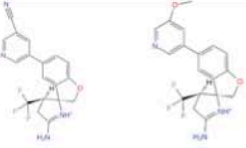
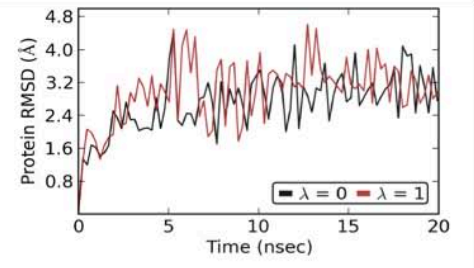
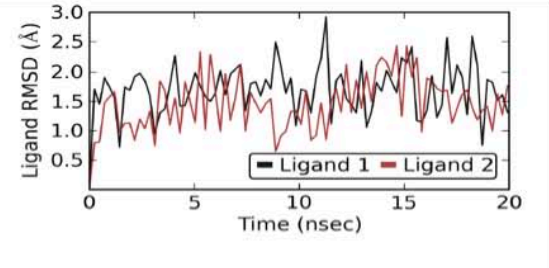

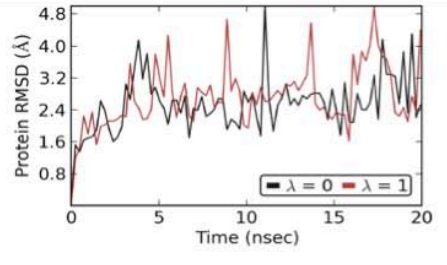
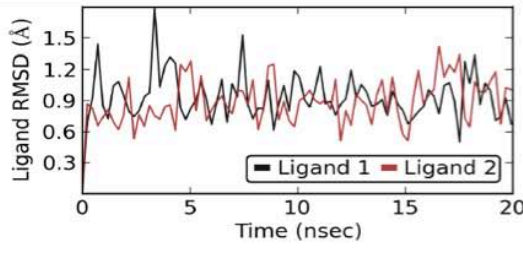
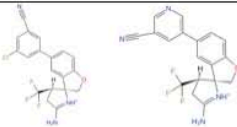
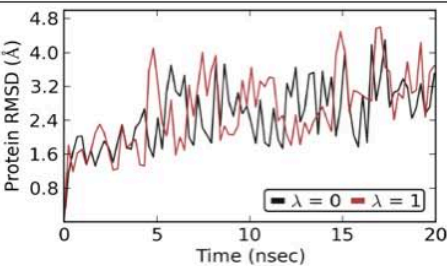
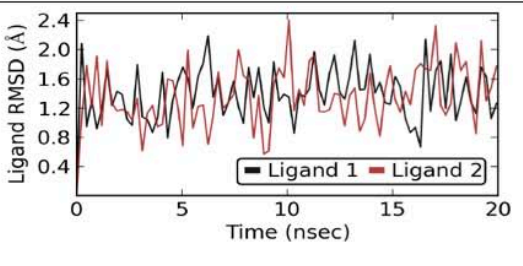
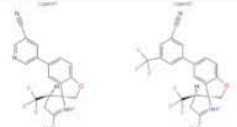
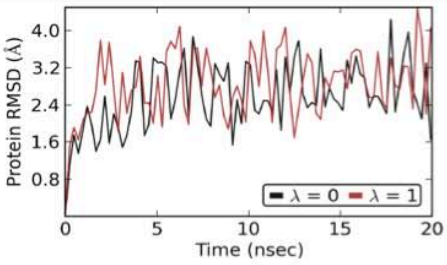
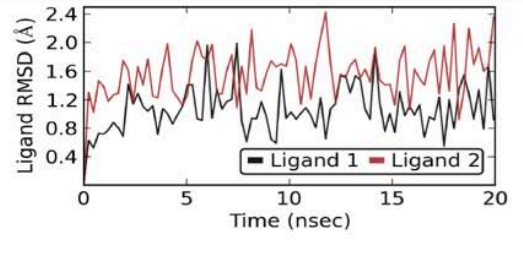
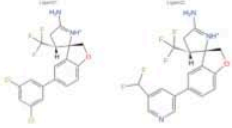
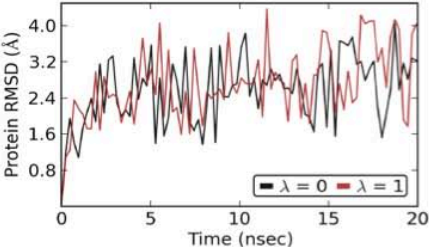
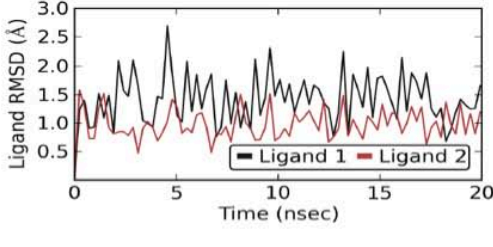

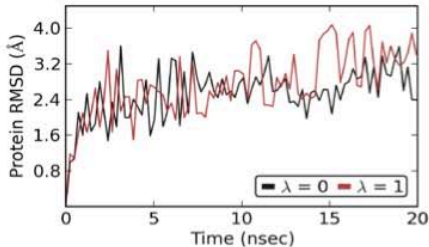
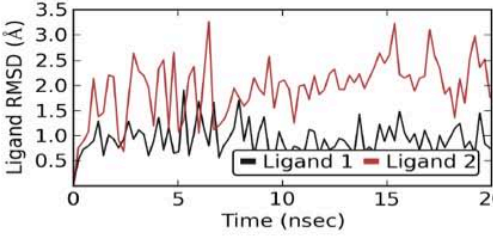

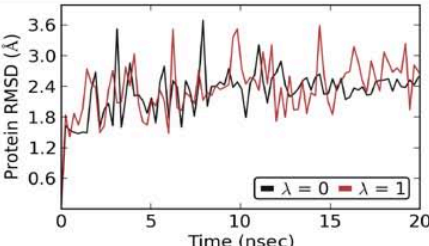
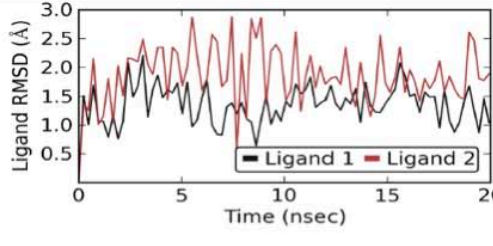


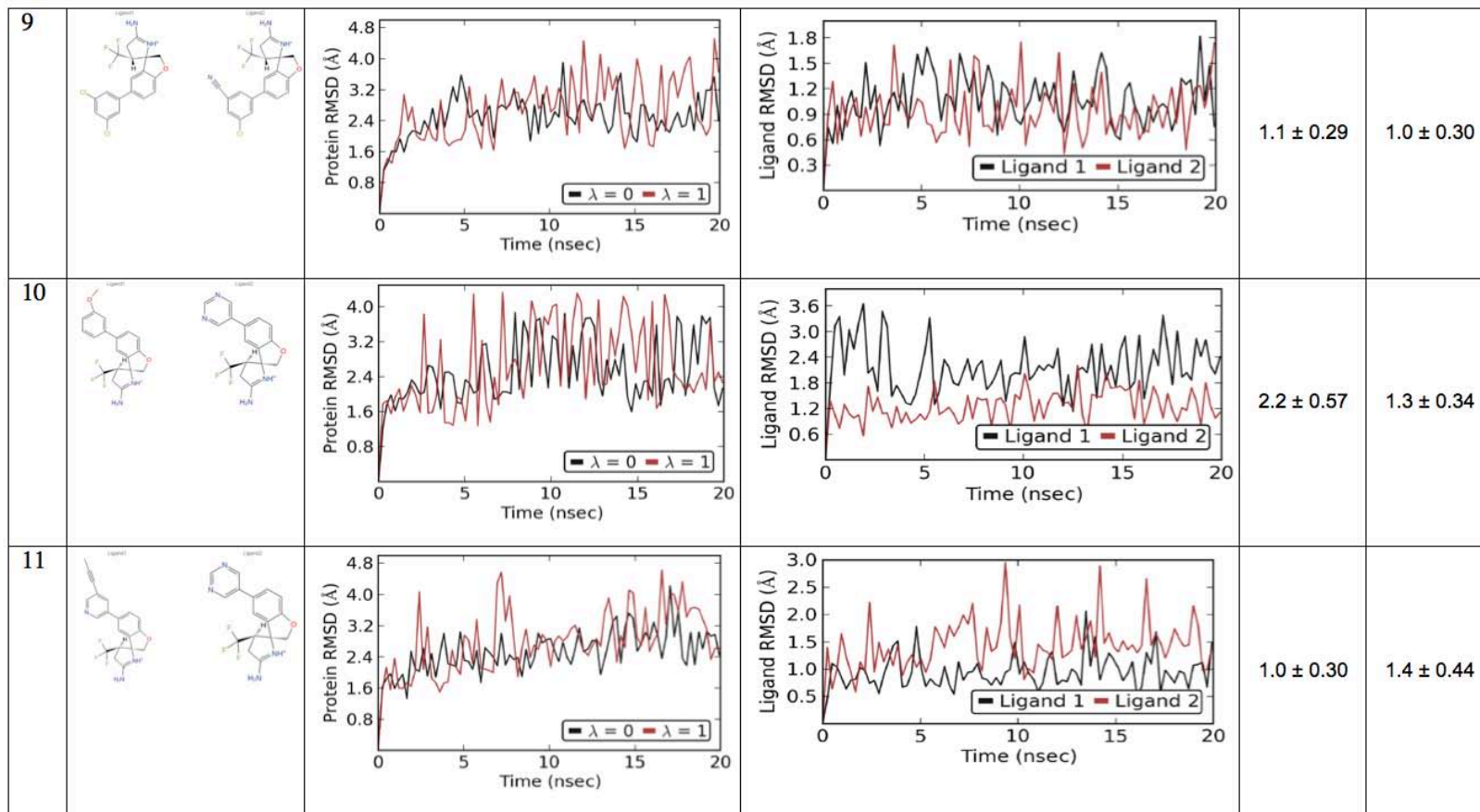
Figure S2. The FEP+ mapper output showing the molecular perturbations performed in the calculations for the prospective application. The numbers on each edge represent the similarity score between the pairs of molecules.


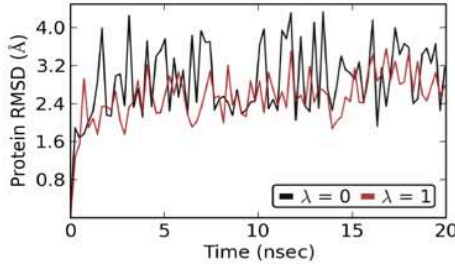
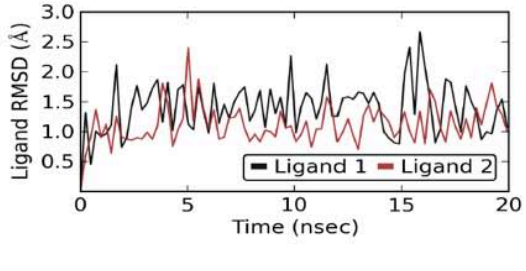
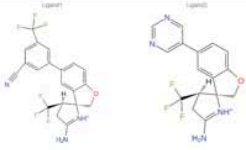
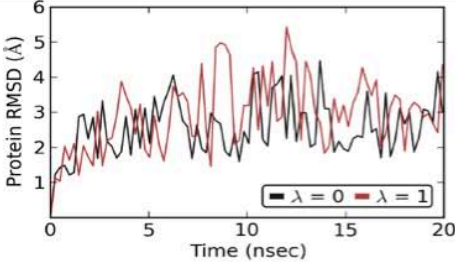
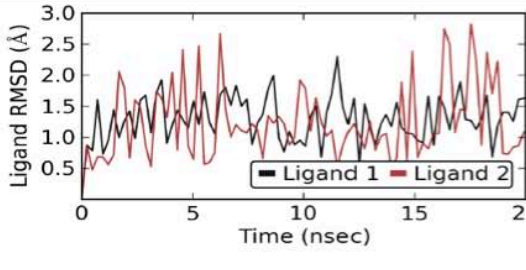

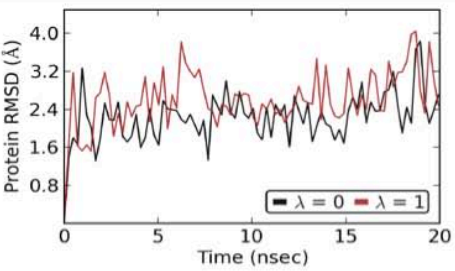
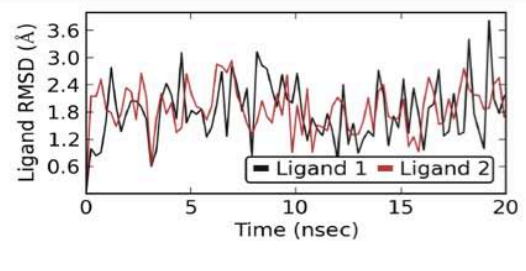
Table S1. Protein back-bone and Ligand RMSD for all perturbations from 20 ns FEP simulations run in the prospective application. RMSD for protein shows the two end points of the perturbation, $\lambda = 0$ and 1. RMSD of both ligands is with respect to the protein.

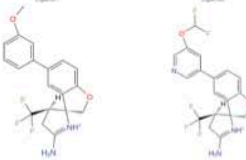
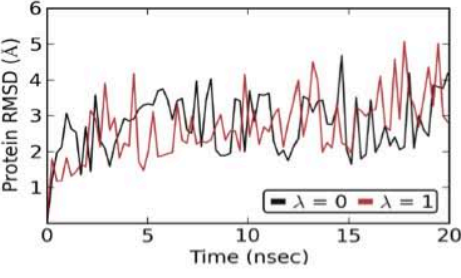
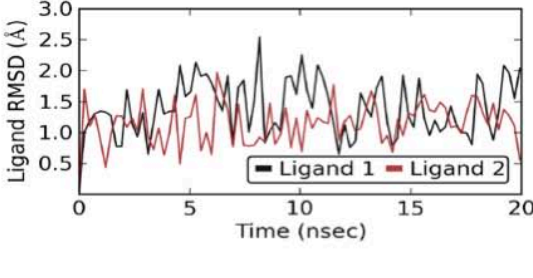
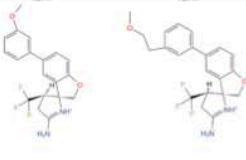
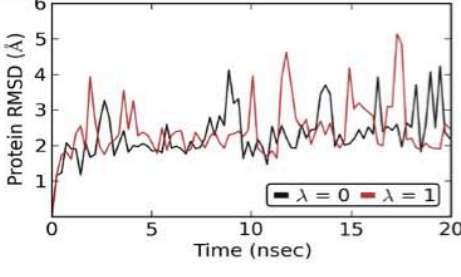
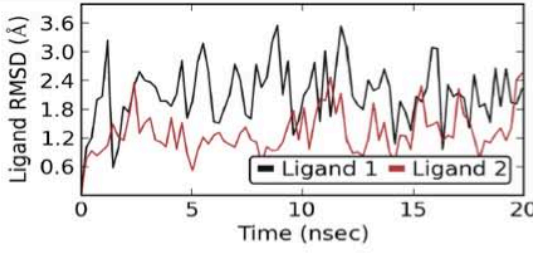
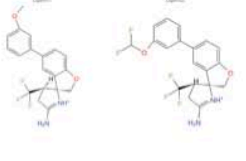
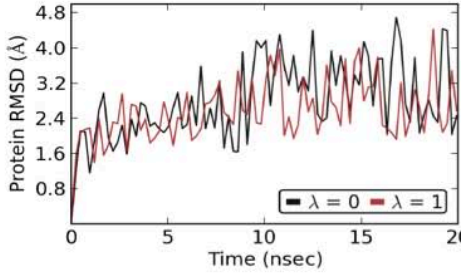
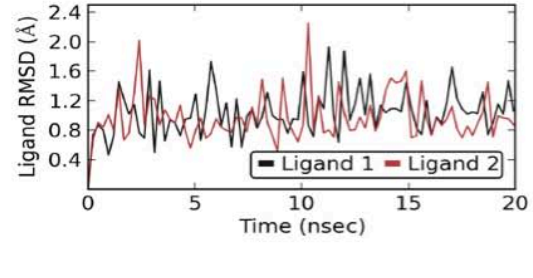
	Ligands 1 and 2	20 ns λ simulation details		RMSD (Å) in complex	
				Ligand 1	Ligand 2
1				1.7 ± 0.55	1.0 ± 0.30
2				1.7 ± 0.40	1.5 ± 0.43

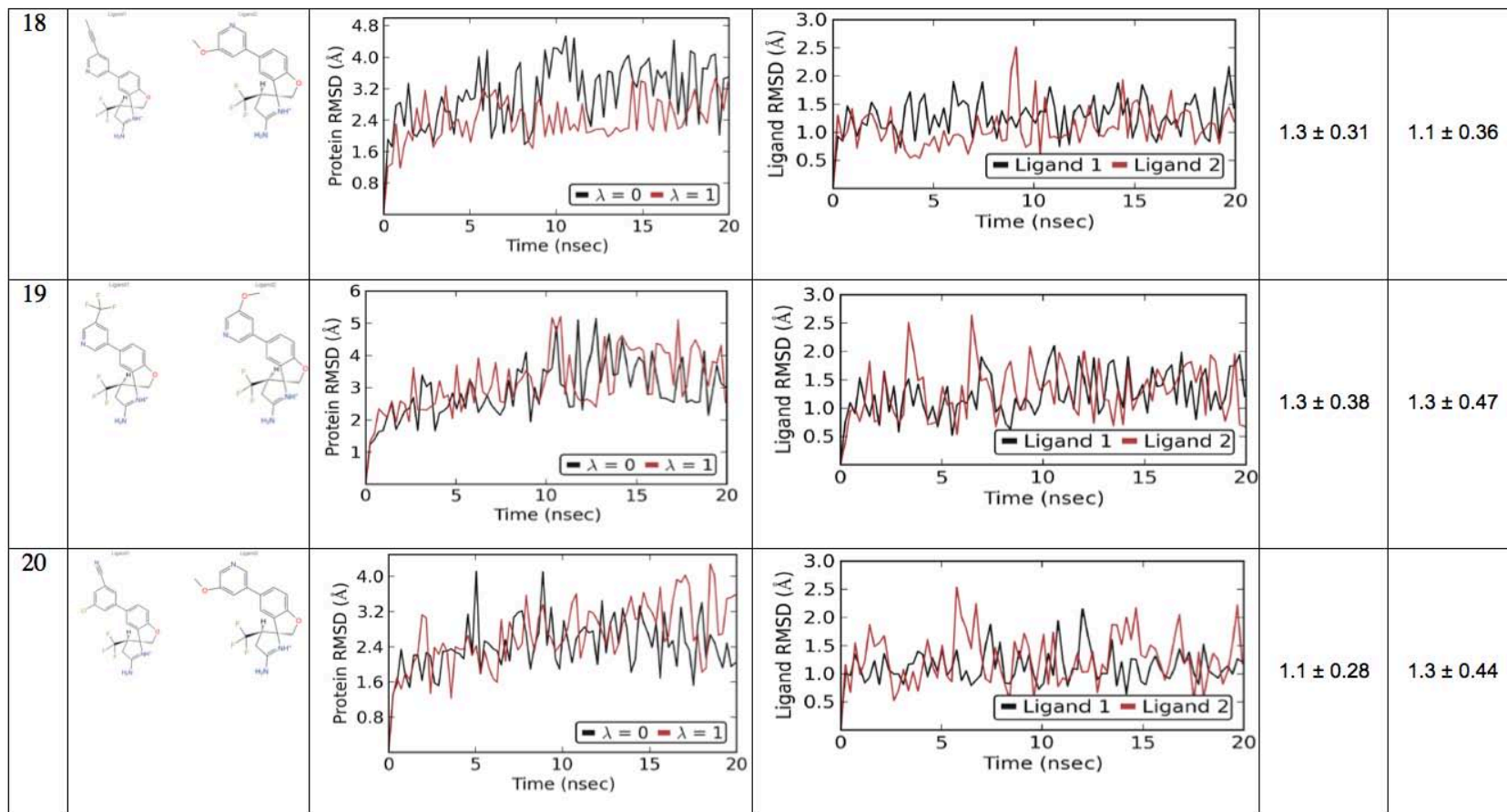
3				0.9 ± 0.21	0.9 ± 0.20
4				1.4 ± 0.34	1.4 ± 0.39
5				1.1 ± 0.32	1.6 ± 0.32

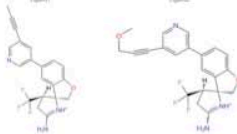
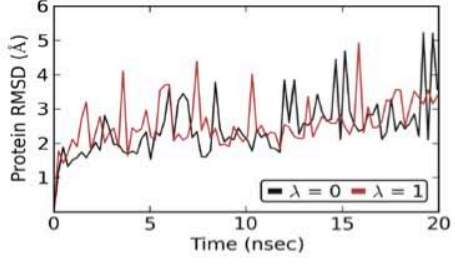
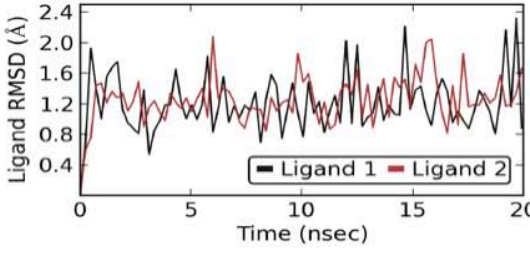
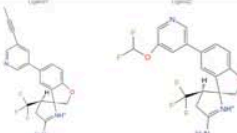
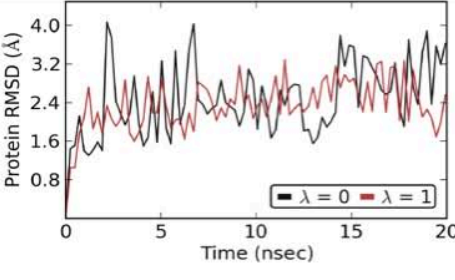
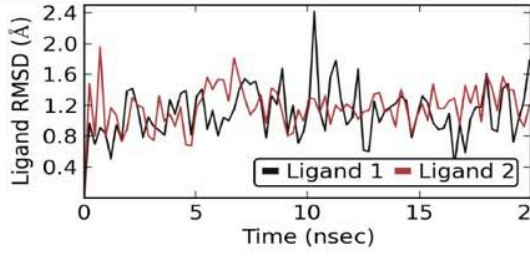
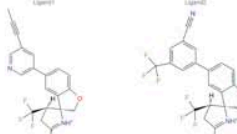
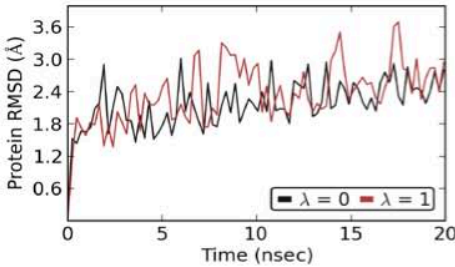
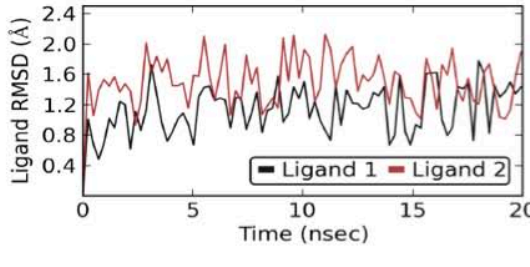
6				1.4 ± 0.41	1.0 ± 0.25
7				0.9 ± 0.30	2.0 ± 0.55
8				1.4 ± 0.32	1.9 ± 0.48



12				1.4 ± 0.41	1.1 ± 0.30
13				1.3 ± 0.35	1.2 ± 0.60
14				1.8 ± 0.66	1.9 ± 0.49

15				1.4 ± 0.43	1.2 ± 0.32
16				2.1 ± 0.60	1.3 ± 0.44
17				1.0 ± 0.30	1.0 ± 0.31



21				1.2 ± 0.36	1.3 ± 0.28
22				1.1 ± 0.33	1.2 ± 0.25
23				1.1 ± 0.30	1.5 ± 0.29

4.7.6. References

- 1 Organization, W. H. Alzheimer's Disease International. World Alzheimer Report. (2009).
- 2 Yiannopoulou, K. G. & Papageorgiou, S. G. Current and future treatments for Alzheimer's disease. *Therapeutic Advances in Neurological Disorders* **6**, 19-33, (2013).
- 3 Hüll, M., Berger, M. & Heneka, M. Disease-Modifying Therapies in Alzheimer's Disease. *Drugs* **66**, 2075-2093, (2006).
- 4 Citron, M. Alzheimer's disease: strategies for disease modification. *Nat Rev Drug Discov* **9**, 387-398 (2010).
- 5 Léger, G. C. & Massoud, F. Novel disease-modifying therapeutics for the treatment of Alzheimer's disease. *Expert Review of Clinical Pharmacology* **6**, 423-442, (2013).
- 6 Hardy, J. A. & Higgins, G. A. Alzheimer's disease: the amyloid cascade hypothesis. *Science* **256**, 184 (1992).
- 7 Karran, E., Mercken, M. & Strooper, B. D. The amyloid cascade hypothesis for Alzheimer's disease: an appraisal for the development of therapeutics. *Nat Rev Drug Discov* **10**, 698-712 (2011).
- 8 Schnabel, J. Amyloid: Little proteins, big clues. *Nature* **475**, S12-S14 (2011).
- 9 Ghosh, A. K., Brindisi, M. & Tang, J. Developing β -secretase inhibitors for treatment of Alzheimer's disease. *Journal of Neurochemistry* **120**, 71-83, (2012).
- 10 Aurn, K. G., Nagaswamy, K. & Jordan, T. Recent Developments of Structure Based β -Secretase Inhibitors for Alzheimers Disease. *Current Topics in Medicinal Chemistry* **5**, 1609-1622, (2005).
- 11 Hong, L. *et al.* Structure of the Protease Domain of Memapsin 2 (β -Secretase) Complexed with Inhibitor. *Science* **290**, 150 (2000).
- 12 Oehlrich, D., Prokopcova, H. & Gijssen, H. J. M. The evolution of amidine-based brain penetrant BACE1 inhibitors. *Bioorganic & Medicinal Chemistry Letters* **24**, 2033-2045, (2014).
- 13 Ghosh, A. K. & Osswald, H. L. BACE1 (β -Secretase) Inhibitors for the Treatment of Alzheimer's Disease. *Chemical Society reviews* **43**, 6765-6813, (2014).
- 14 Edwards, P. D. *et al.* Application of Fragment-Based Lead Generation to the Discovery of Novel, Cyclic Amidine β -

- Secretase Inhibitors with Nanomolar Potency, Cellular Activity, and High Ligand Efficiency. *Journal of Medicinal Chemistry* **50**, 5912-5925, (2007).
- 15 Baxter, E. W. *et al.* 2-Amino-3,4-dihydroquinazolines as Inhibitors of BACE-1 (β -Site APP Cleaving Enzyme): Use of Structure Based Design to Convert a Micromolar Hit into a Nanomolar Lead. *Journal of Medicinal Chemistry* **50**, 4261-4264, (2007).
- 16 Tresadern, G. *et al.* Rational design and synthesis of aminopiperazinones as β -secretase (BACE) inhibitors. *Bioorganic & Medicinal Chemistry Letters* **21**, 7255-7260, (2011).
- 17 Fustero, S. *et al.* Design, Synthesis, and Biological Evaluation of Novel Fluorinated Ethanolamines. *Chemistry – A European Journal* **17**, 14772-14784, (2011).
- 18 Mateu, N. *et al.* A Versatile Approach to CF₃-Containing 2-Pyrrolidones by Tandem Michael Addition–Cyclization: Exemplification in the Synthesis of Amidine Class BACE1 Inhibitors. *Chemistry – A European Journal* **21**, 11719-11726, (2015).
- 19 Rombouts, F. J. R. *et al.* 1,4-Oxazine β -Secretase 1 (BACE1) Inhibitors: From Hit Generation to Orally Bioavailable Brain Penetrant Leads. *Journal of Medicinal Chemistry* **58**, 8216-8235, (2015).
- 20 Malamas, M. S. *et al.* Aminoimidazoles as Potent and Selective Human β -Secretase (BACE1) Inhibitors. *Journal of Medicinal Chemistry* **52**, 6314-6323, (2009).
- 21 Malamas, M. S. *et al.* Design and Synthesis of 5,5'-Disubstituted Aminohydantoin as Potent and Selective Human β -Secretase (BACE1) Inhibitors. *Journal of Medicinal Chemistry* **53**, 1146-1158, (2010).
- 22 Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Accounts of Chemical Research* **42**, 724-733, (2009).
- 23 *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Vol. 86 (Springer Series in Chemical Physics, 2007).
- 24 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977).
- 25 Jorgensen, W. L. & Ravimohan, C. Monte Carlo simulation of differences in free energies of hydration. *The Journal of Chemical Physics* **83**, 3050-3054, (1985).

- 26 Bash, P. A., Singh, U. C., Langridge, R. & Kollman, P. A. Free energy calculations by computer simulation. *Science* **236**, 564 (1987).
- 27 Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews* **93**, 2395-2417, (1993).
- 28 Wong, C. F. & McCammon, J. A. Dynamics and design of enzymes and inhibitors. *Journal of the American Chemical Society* **108**, 3830-3832, (1986).
- 29 Merz, K. M. & Kollman, P. A. Free energy perturbation simulations of the inhibition of thermolysin: prediction of the free energy of binding of a new inhibitor. *Journal of the American Chemical Society* **111**, 5649-5658, (1989).
- 30 Essex, J. W., Severance, D. L., Tirado-Rives, J. & Jorgensen, W. L. Monte Carlo Simulations for Proteins: Binding Affinities for Trypsin–Benzamidine Complexes via Free-Energy Perturbations. *The Journal of Physical Chemistry B* **101**, 9663-9669, (1997).
- 31 Deng, Y. & Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **113**, 2234-2246, (2009).
- 32 Durrant, J. D. & McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **9**, 71-71, (2011).
- 33 Michel, J. & Essex, J. W. Hit Identification and Binding Mode Predictions by Rigorous Free Energy Simulations. *Journal of Medicinal Chemistry* **51**, 6654-6664, (2008).
- 34 Wang, L. *et al.* Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **137**, 2695-2703, (2015).
- 35 Rombouts, F. J. R. *et al.* Pyrido[4,3-e][1,2,4]triazolo[4,3-a]pyrazines as Selective, Brain Penetrant Phosphodiesterase 2 (PDE2) Inhibitors. *ACS Medicinal Chemistry Letters* **6**, 282-286, (2015).
- 36 Barman, A., Schürer, S. & Prabakar, R. Computational Modeling of Substrate Specificity and Catalysis of the β -Secretase (BACE1) Enzyme. *Biochemistry* **50**, 4337-4349, (2011).
- 37 Polgár, T. & Keserü, G. M. Virtual Screening for β -Secretase (BACE1) Inhibitors Reveals the Importance of Protonation States

- at Asp32 and Asp228. *Journal of Medicinal Chemistry* **48**, 3749-3755, (2005).
- 38 Ellis, C. R. & Shen, J. pH-Dependent Population Shift Regulates BACE1 Activity and Inhibition. *Journal of the American Chemical Society* **137**, 9543-9546, (2015).
- 39 Craig, I. R., Essex, J. W. & Spiegel, K. Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *Journal of Chemical Information and Modeling* **50**, 511-524, (2010).
- 40 Roos, K., Viklund, J., Mueller, J., Kaspersson, K. & Svensson, M. Potency Prediction of β -Secretase (BACE-1) Inhibitors Using Density Functional Methods. *Journal of Chemical Information and Modeling* **54**, 818-825, (2014).
- 41 Bukhtiyarov, Y. C., S.; Dillard, L. W.; Dorner-Ciossek, C.; Fuchs, K.; Jia, L.; Lala, D. S.; Morales-Ramos, A.; Rast, G.; Reeves, J.; Singh, S. B.; Venkatraman, S.; Xu, Z.; Yuan, J.; Zhao, Y.; Zheng, Y. Spirocyclic Acylguanidine Derivatives as β -Secretase Inhibitors Useful in Treatment of Diseases Characterized by Production of β -Amyloid Aggregates and Their Prevention. (2013).
- 42 Thomas, A. A. *et al.* 8-Tetrahydropyran-2-yl Chromans: Highly Selective Beta-Site Amyloid Precursor Protein Cleaving Enzyme 1 (BACE1) Inhibitors. *Journal of Medicinal Chemistry* **57**, 10112-10129, (2014).
- 43 Hunt, K. W. *et al.* Spirocyclic β -Site Amyloid Precursor Protein Cleaving Enzyme 1 (BACE1) Inhibitors: From Hit to Lowering of Cerebrospinal Fluid (CSF) Amyloid β in a Higher Species. *Journal of Medicinal Chemistry* **56**, 3379-3403, (2013).
- 44 Csjernyik, G. K., S.; Kers, A.; Kolmodin, K.; Nylöf, M.; Ohberg, L.; Rakos, L.; Ohberg, L.; Rakos, L.; Sandberg, L.; Sehgelmeble, F.; Söderman, P.; Schwahn, B.M.; Von Berg, S. Derivatives and Analogs for use as BACE Inhibitors. (2012).
- 45 Bohlin, M. H. S., C. R. Preparation of Camsylate Salt of (1r, 1'R,4R)-4-methoxy-5''-methyl-6'-[5-(prop-1-yn-1-yl)pyridine-3-yl]-3'H-dispiro[cyclohexane-1,2'-inden-1,2'-imidazole]-4''-amine Useful for Treating A β -Related Pathologies. (2013).
- 46 Xu, Y. *et al.* Flexibility of the flap in the active site of BACE1 as revealed by crystal structures and molecular dynamics simulations. *Acta Crystallographica Section D* **68**, 13-25, (2012).

- 47 Banner, D. W. *et al.* Mapping the conformational space accessible to BACE2 using surface mutants and co-crystals with Fab fragments, Fynomers and Xaperones. *Acta Crystallographica Section D* **69**, 1124-1137, (2013).
- 48 Liu, S. *et al.* Lead optimization mapper: automating free energy calculations for lead optimization. *Journal of Computer-Aided Molecular Design* **27**, 755-770, (2013).
- 49 We routinely add additional perturbations to the Mapper. For comparison we performed the 5 ns FEP+ simulation using the default mapper without extra connections. In this case, results were similar: the cycle closure errors ranged from 0.11 to 0.70 kcal/mol with a mean of 0.45 kcal/mol, the FEP+ predicted error had a maximum of 1.25 and mean of 0.91 kcal/mol, and ultimately an MUE versus experiment of 0.82 kcal/mol was found. Hence, in this case the dataset is relatively insensitive to the addition of extra perturbations into the Mapper.
- 50 The similarity score failed to calculate for one perturbation between molecules labelled (3S,3'R)-17 and (3S,3'R)-8, but this was a relatively small perturbation involving loss of the cyano group and changing an sp² carbon to nitrogen to form the pyridyl ring.
- 51 (-)-(3R,3'S) enantiomers (all levorotatory in these series) were inactive (IC₅₀ > 10 μM in cellular and enzymatic assays).
- 52 Li, Y., Han, L., Liu, Z. & Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *Journal of Chemical Information and Modeling* **54**, 1717-1736, (2014).
- 53 Steinbrecher, T. B. *et al.* Accurate Binding Free Energy Predictions in Fragment Optimization. *Journal of Chemical Information and Modeling* **55**, 2411-2420, (2015).
- 54 Christ, C. D. & Fox, T. Accuracy Assessment and Automation of Free Energy Calculations for Drug Design. *Journal of Chemical Information and Modeling* **54**, 108-120, (2014).
- 55 Lovering, F. *et al.* Imidazotriazines: Spleen Tyrosine Kinase (Syk) Inhibitors Identified by Free-Energy Perturbation (FEP). *ChemMedChem* **11**, 217-233, (2016).
- 56 Bollini, M. *et al.* Computationally-Guided Optimization of a Docking Hit to Yield Catechol Diethers as Potent Anti-HIV Agents. *Journal of Medicinal Chemistry* **54**, 8582-8591, (2011).

- 57 Shoichet, B. K., Walters, W. P., Jiang, H. & Bajorath, J. Advances in Computational Medicinal Chemistry: A Reflection on the Evolution of the Field and Perspective Going Forward. *Journal of Medicinal Chemistry* **59**, 4033-4034, (2016).
- 58 Harder, E. *et al.* OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *Journal of Chemical Theory and Computation* **12**, 281-296, (2016).
- 59 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Research* **28**, 235-242 (2000).
- 60 Maestro (120 West 45th Street 17th Floor, Tower 45 New York, NY 10036-4041, USA, 2014).
- 61 Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **27**, 221-234, (2013).
- 62 Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *Journal of Chemical Theory and Computation* **7**, 2284-2295, (2011).
- 63 Watts, K. S. *et al.* ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers. *Journal of Chemical Information and Modeling* **50**, 534-546, (2010).
- 64 Shivakumar, D., Harder, E., Damm, W., Friesner, R. A. & Sherman, W. Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *Journal of Chemical Theory and Computation* **8**, 2553-2558, (2012).
- 65 Bowers, K. J. C., E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. *in In Proceedings of the 2006 ACM/IEEE conference on Supercomputing*.
- 66 Liu, P., Kim, B., Friesner, R. A. & Berne, B. J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13749-13754 (2005).
- 67 Wang, L., Friesner, R. A. & Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange

- with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **115**, 9431-9438, (2011).
- 68 Wang, L. *et al.* Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. *Journal of Chemical Theory and Computation* **9**, 1282-1293, (2013).
- 69 Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **22**, 245-268, (1976).
- 70 Paliwal, H. & Shirts, M. R. A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods. *Journal of Chemical Theory and Computation* **7**, 4115-4134, (2011).

5. Conclusions

5. Conclusions

General conclusions

This work offers further confidence for the important role that computation and simulation can have in furthering our understanding of biology and chemistry problems.

Overall, in this thesis state of the art molecular dynamics applications have contributed understanding of biological phenomena, such as, the mechanism of action of a complex GPCR heterotetramer, the allosteric modulation of class C GPCRs, and the more accurate prediction of binding energies for drug discovery

5.1. The quaternary structure of the adenosine A₁-A_{2A} receptor heteromer

Using computer modeling, aided by bioluminescence resonance energy transfer assays to monitor receptor homomerization and heteromerization and G-protein coupling, we predicted the interacting interfaces of the A₁-A_{2A}Het and propose a quaternary structure of the GPCR tetramer in complex with two different G proteins (Gi and Gs). The molecular architecture consists of a rhombus-shaped heterotetramer, which is bound to two interacting heterotrimeric G proteins in the external protomers.

These novel results constitute an important advance in understanding the molecular intricacies involved in GPCR function.

5.2. Allosteric modulators of class C GPCRs

I have performed a combined experimental and computational study to reveal for the first time that allosteric modulators of mGlu₂ receptors interact with the homologous ‘trigger switch’ and ‘transmission switch’ amino acids as seen in class A GPCRs. I have built pharmacophore models for positive (PAMs) and negative (NAMs) allosteric modulators and proposed the binding modes of 3 PAMs and 3 NAMs using SAR and mutagenesis data provided by Janssen. Furthermore, I have evaluated the proposed docking models by MD simulations. Analysis of the MD trajectories has provided key information about the modulation of class C GPCRs by either PAMs or NAMs concluding that this modulation involves rearrangement of homologous ‘switches’ as (in)activation of class A by either orthosteric antagonists or agonists.

5.3. Design selective beta-secretase-1 inhibitors

Free energy perturbation is an important computational tool for predicting relative binding energies between ligands and protein targets.

Novel s piroaminodihydropyrroles pr obing for opt imized interactions at t he P 3 p ocket o f beta-secretase-1 (B ACE1) w ere designed w ith t he use o f FEP calculations. A set o f 18 f unctional groups t argeting t he P 3 p ocket o f BACE1 w ere s elected f or calculations. B ased on t he F EP pr edictions, 9 c ompounds w ere selected f or synthesis and pharmacological evaluation. The quality o f t he pr ediction be nefited f rom i ncreased s imulation t ime. T he selected substituents approach t he 10s loop o f t he P3 pocket. This is known t o be a r egion o f c onformational f lexibility, hence, t he protein may also need t o adapt t o t his particular group and may do so on a slow timescale t hat benefits f rom extra sampling. Additional retrospective application w as a lso s tudied i n t his w ork. The retrospective application s tudied 32 m olecules a nd s howed good correlation b etween p redicted and e xperimental bi nding energies, which w as largely consistent w ith respect t o s imulation t ime. Our simulations s howed good c orrelation b etween pr edicted and experimental binding energies.

The F EP m ethod ou tperformed do cking a nd M M-GBSA approaches i n both retrospective and prospective applications.

6. List of publications

1. Cid, J. M.; Tresadern, G.; Vega, J. A.; de Lucas, A. I.; del Cerro, A.; Matesanz, E.; Linares, M. L.; García, A.; Iturrino, L.; Pérez-Benito, L.; Macdonald, G. J.; Oehlrich, D.; Lavreysen, H.; Peeters, L.; Ceusters, M.; Ahnaou, A.; Drinkenburg, W.; Mackie, C.; Somers, M.; Trabanco, A. A., Discovery of 8-Trifluoromethyl-3-cyclopropylmethyl-7-[(4-(2,4-difluorophenyl)-1-piperazinyl)methyl]-1,2,4-triazolo[4,3-a]pyridine (JNJ-46356479), a Selective and Orally Bioavailable mGlu2 Receptor Positive Allosteric Modulator (PAM). *Journal of Medicinal Chemistry* **2016**, *59*, 8495-8507.

2. Ciordia, M.; Pérez-Benito, L.; Delgado, F.; Trabanco, A. A.; Tresadern, G., Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. *Journal of Chemical Information and Modeling* **2016**, *56*, 1856-1871.

3. Navarro, G.; Cordoní, A.; Zelman-Femiak, M.; Brugarolas, M.; Moreno, E.; Aguinaga, D.; Pérez-Benito, L.; Cortés, A.; Casadó, V.; Mallol, J.; Canela, E. I.; Lluís, C.; Parado, L.; García-Sáez, A. J.; McCormick, P. J.; Franco, R., Quaternary structure of a G-protein-coupled receptor heterotetramer in complex with G(i) and G(s). *BMC Biology* **2016**, *14*, 26.

4. Doornbos, M. L. J.; Pérez-Benito, L.; Tresadern, G.; Mulder-Krieger, T.; Biesmans, I.; Trabanco, A. A.; Cid, J. M.; Lavreysen, H.; Ijzerman, A. P.; Huitman, L. H., Molecular mechanism of positive allosteric modulation of the metabotropic glutamate receptor 2 by JNJ-46281222. *British Journal of Pharmacology* **2016**, *173*, 588-600.

5. Mateu, N.; Ciordia, M.; Delgado, O.; Sánchez-Roselló, M.; Trabanco, A. A.; Van Gool, M.; Tresadern, G.; Pérez-Benito, L.; Fustero, S., A Versatile Approach to CF₃-Containing 2-Pyrrolidones by Tandem Michael Addition–Cyclization: Exemplification in the Synthesis of Amidine Class BACE1 Inhibitors. *Chemistry – A European Journal* **2015**, *21*, 11719-11726.

6. Masià-Balagué, M.; Izquierdo, I.; Garrido, G.; Cordoní, A.; Pérez-Benito, L.; Miller, N. L. G.; Schlaepfer, D. D.; Gigoux, V.; Aragay, A. M., Gastrin-stimulated G α (13) Activation of Rgnef Protein (ArhGEF28) in LDL-1 Colon Carcinoma Cells. *The Journal of Biological Chemistry* **2015**, *290*, 15197-15209.

Manuscripts

1. Henrik Keränen, Laura Pérez-Benito, Myriam Ciordia, Francisca Delgado, Thomas Steinbrecher, Daniel Oehlich, Herman van Vlijmen, Andrés A. Trabanco, Gary Tresadern. Acylguanidine Beta Secretase Inhibitors: A Combined Experimental and Free Energy Perturbation Study. *Manuscript under review at J. Chem. Theory. Comput.*
 2. Laura Pérez-Benito, Doornbos, M. L. J., Arnau C ordomí, Luc Peeters, Hilde Lavreysen, Gary Tresadern, Leonardo Pardo The transmission switch mechanism of allosteric modulation of the metabotropic glutamate 2 receptor. *Manuscript under review at Proc. Natl. Acad. Sci.*
 3. Gemma Navarro, Arnau C ordomí, Marc Bugarolas, Estefanía Moreno, David Aguinaga, Laura Pérez-Benito, Sergi Ferrer, Antoni Cortés, Vicent Casadó, Josefa Mallol, Enric I. Canela, Carme Lluís, Leonardo Pardo, Peter J. McCormick and Rafael Franco. The C-terminal end of a GPCR enables allosteric communications between G_i and G_s in a G-protein-coupled receptor heteromer. *Manuscript under preparation*
 4. Laura Pérez-Benito, Andrew Henry, Minos-Timotheos Matsoukas, Arnau C ordomí, Gary Tresadern and Leonardo Pardo. Bivalent ligands: The size matters. Computational Design and Analysis of GPCR Bivalent Ligands. *Manuscript under preparation*
 5. Laura Pérez-Benito, Eduardo Mayol, Mireia Jiménez-Rosés. Arnau C ordomí and Leonardo Pardo. Assessment of force fields for membrane protein simulation. *Manuscript under preparation*
-

