Implementation of an original workflow for allosteric drug design based on state-based pharmacology with the accurate calculation of the protein-ligand binding affinity for a defined conformation of the receptor based on HPC resources (ChemFlow, work in progress) (D5.10 - SGA3)

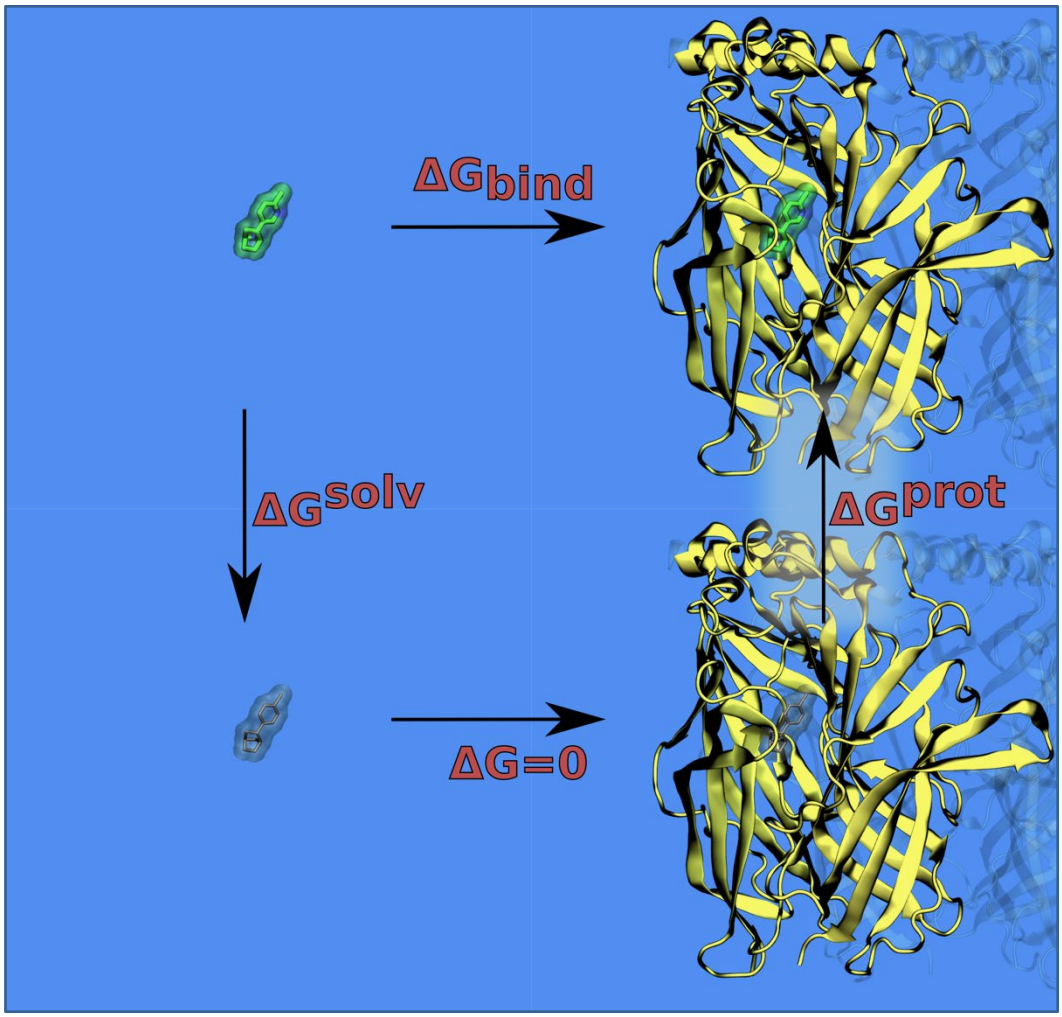

Figure 1: A prototype for accurate computational neuropharmacology.

Project Number:	945539	Project Title:	HBP SGA3

Document Title:	Implementation of an original workflow for allosteric drug design based on state-based pharmacology with the accurate calculation of the protein-ligand binding affinity for a defined conformation of the receptor based on HPC resources (ChemFlow, work in progress).
Document Filename:	D5.10 (D86) SGA3 M24 SUBMITTED 220425.docx
Deliverable Number:	SGA3 D5.10 (D86)
Deliverable Type:	Other
Dissemination Level:	PU = Public
Planned Delivery Date:	SGA3 M24 / 31 Mar 2022
Actual Delivery Date:	SGA3 M25 / 24 Apr 2022 (resubmitted 05 Dec 2023
Author(s):	Jean-Pierre CHANGEUX, IP (P22); Marco CECCHINI, University of Strasbourg; Adrien CERDAN, IP (P22) and University of Strasbourg
Compiled by:	Marco CECCHINI, University of Strasbourg Adrien CERDAN, IP (P22) and University of Strasbourg
Contributor(s):	
WP QC Review:	Evita MAILLI, ATHENA (P133)
WP Leader / Deputy Leader Sign Off:	Yannis IOANNIDIS, ATHENA (P133)
T7.4 QC Review:	Martin TELEFONT, Annemieke MICHELS, EBRAINS (P1)
Description in GA:	ChemFlow aims at preparing, running, and analysing a virtual screening campaign based on docking with free energy rescoring, efficiently and semi-automatically. ChemFlow is currently tested on a number of protein-ligand complexes and shall be distributed freely within the end of SGA.
Abstract:	The extension of the MWC model of allostery provides a link between pharmacological properties such as potency, efficacy, and selectivity, to state-dependant binding affinities. This framework is well suited to study the modulation of pentameric Ligand-Gated ion Channels (pLGICs) through computational techniques, such as Molecular Dynamics and Free-Energy Perturbation, which provides methods for binding affinity calculations. Here we report on the application of Alchemical (i.e., non-physical) transformations for the accurate calculations of binding affinity of small-molecules in the acetylcholine-binding protein (AChBP), a model system for the nicotinic acetylcholine Receptor (nAChR). The development of accurate ligand-binding free energy calculations based on Molecular Dynamics simulations and high-resolution structures of pLGICs is a critical step towards the implementation of effective numerical strategies powered by HPC resources (e.g. ChemFlow) for allosteric drug design in the brain.
Keywords:	pentameric ligand-gated ion channel, allosteric drug design, binding affinity, free energy perturbation, AChBP, nAChR, Molecular Dynamics, HPC resources, virtual screening (ChemFlow)
Target Users/Readers:	Researchers, Computational chemists

Table of Contents

Table of Figures

Date	Change Requested / Change Made / Other Action
25/04/2022	Deliverable submitted to EC
02/12/2023	Resubmission with specified changes requested in Review Report Main changes requested: • Change 1 (extract from Review Report) *"The deliverable nicely describes a science case, but its relevance to EBRAINS is unclear. It needs to be revised, explaining how EBRAINS RI was used in order to achieve the described results."* • Change 2 (extract from Review Report) *"In spite of the important goal, the work described in D5.10 won't be completed. Minimal justification is given in the text besides the lack of resources. From the Section 6 of the deliverable, it looks like the failed implementation was due to an initial wrong guess of hypothesis. This should be confirmed."* • Change 3 (extract from Review Report) *"The deliverable works seems to be linked to task T5.19 but not clear enough."*
	Revised draft sent by WP to PCO. Main changes made, with indication where each change was made: • Change 1 The deliverable aims to predict the mode of action of neurotransmitters and the design of drugs targeting synaptic receptors for the creation of new treatments. To this aim, we developed a theoretical framework based on the MWC model to predict the pharmacological attributes of potency, efficacy and selectivity of the modulatory ligands (10.1016/j.mam.2021.101044) and attempted its implementation via rigorous binding free energy calculations based on Molecular Dynamics. The original goals were providing a proof-of-principle of this approach and providing an implementation on EBRAINS. Since the goal was not achieved no implementation is currently available from EBRAINS. • Change 2 (see Section 6)

	The work could not be completed because it was found that absolute binding free energy calculations based on classical MD on the ligand-binding domain of the nicotinic receptor are not accurate enough (i.e. poor correlation with experiments). The initial hypothesis on the design of agonists, antagonists, and allosteric modulators was not confuted. Rather its implementation turned out to be more challenging than expected pointing to shortcomings both in state-of-art ligand-binding free energy calculations or models of protein-ligand interactions for MD (i.e. force fields). • Change 3 This deliverable had minimal effect on task T5.19 since modulatory compounds tested for COVID19 were extracted from the existing literature (ACRALL, https://doi.org/10.25493/7BWK-RM5) and not rationally designed.
	Revised version resubmitted to EC by PCO via SyGMa

1. Introduction

A fundamental ethical and societal issue is the overall burden of brain pathologies in Europe, especially in the Covid-19 era. Neurotransmitter receptors at chemical synapses, like the excitatory nicotinic acetylcholine receptor (nAChR) or the inhibitory GABA receptor (GABAR), are relevant pharmacological targets for several neurological disorders in humans, including Alzheimer's, Parkinson's, schizophrenia, depression, and neuropathic pain. Recently, nAChRs have been implicated in the pathophysiology of the Covid-19 infection.[1] Administration of the anthelmintic ivermectin, which is a positive allosteric modulator of nAChR α7, was shown to provide a protective effect against Covid-19 in a hamster model.[2] The discovery of new therapeutics for brain disorders is a thus critical, yet an unsolved problem with a strong societal impact. During SGA3, we proposed a straightforward extension of the MWC model for neurotransmitter receptor function to account for allosteric regulation and desensitisation.[3] Our theoretical development provides expressions for the pharmacological attributes of the modulatory ligands (i.e., potency, efficacy, and selectivity) in terms of their ligand-binding affinity for the resting, active and desensitised state of the receptor. Although conformation-dependent ligand-binding affinities are hardly accessible experimentally, they are in principle accessible by modelling and simulations. In this context, the development and the implementation of accurate and efficient computational strategies to access ligand-binding free energies is a critical step towards the rational design of neuroactive compounds. The latter will open the field of computational neuropharmacology.

The work reported here belongs to the general perspective of understanding the molecular contributions to the overall simulation of the brain carried out by the HBP. In this context, understanding the mode of action of neurotransmitters and the design of drugs targeting these receptors are critical tasks. Since neurotransmitter receptors were found to exist under a discrete number of conformations with different functional properties (i.e., a resting state with a closed channel that binds antagonists, an active state with an open channel which selectively binds activators or agonists and a desensitised state with a closed channel but high affinity for variety of drugs), there should be a well-defined pharmacology for each of these conformational states, which we refer to it as "state-based pharmacology". The aim of our team in SGA3 and within the HBP perspective is to develop such pharmacology through the understanding of signal transduction in the brain using molecular dynamics simulations and the design of drugs targeting each of these conformational states, i.e., agonists, antagonists or desensitisers, via binding free energy calculations.

This Deliverable aimed at the development of HPC strategies for allosteric drug design in the brain based on high-resolution structures of neurotransmitter receptors and Molecular Dynamics simulations. This has not been done by pharmaceutical companies yet but if achieved will be a real breakthrough in the discipline of drug design. At this stage and due to several difficulties as we shall see, the work could not be completed and no software for accurate binding free energy calculations in pentameric ligand-gated ion channels could be integrated into EBRAINS. The current report deals with this attempt and presents working solutions and protocols[a] to problems encountered on the way towards this very ambitious goal, i.e., the quantitative evaluation of the intrinsic affinity of ligands to particular a conformation of the receptor from first principles. Due to limited funding received in SGA3, we are currently unable to hire someone else to complete the work by M36.

2. General Methodology

The most natural way to evaluate absolute binding free energy in the various states of a neurotransmitter receptor would be by Free-Energy Perturbation (FEP) calculations. This kind of calculation can be carried out by two Molecular Dynamics (MD) techniques, either a geometric pulling of the ligand from its binding site or by an alchemical transformation. Alchemical transformation is referring to the process of turning atoms or molecules into non-physical (i.e., alchemical) species (see Figure 2). Computationally speaking it is often advantageous to consider these alchemical

[a] https://github.com/AdrienCerdan/FEP-auto

intermediate states instead of the physical ones for which the time scale required are out-of-reach from simulations. Since free-energy is a state variable (i.e., independent from the path), one gets meaningful results even using non-physical intermediates. Complete review and comparison of the computational techniques for absolute binding free-energy calculations can be found elsewhere.

We decided to explore the use of equilibrium alchemical transformations for binding since it is most appropriate for binding sites that are buried and for which no straightforward puling coordinates can be defined. The procedure consists of creating a thermodynamic cycle to compute the binding free energy by displacing the ligand from the solution to the binding site of the protein (see Figure 2). Within this cycle we introduce two non-physical transformations named alchemical where the ligand is uncoupled from the water solution first, and then recoupled in the protein binding site.

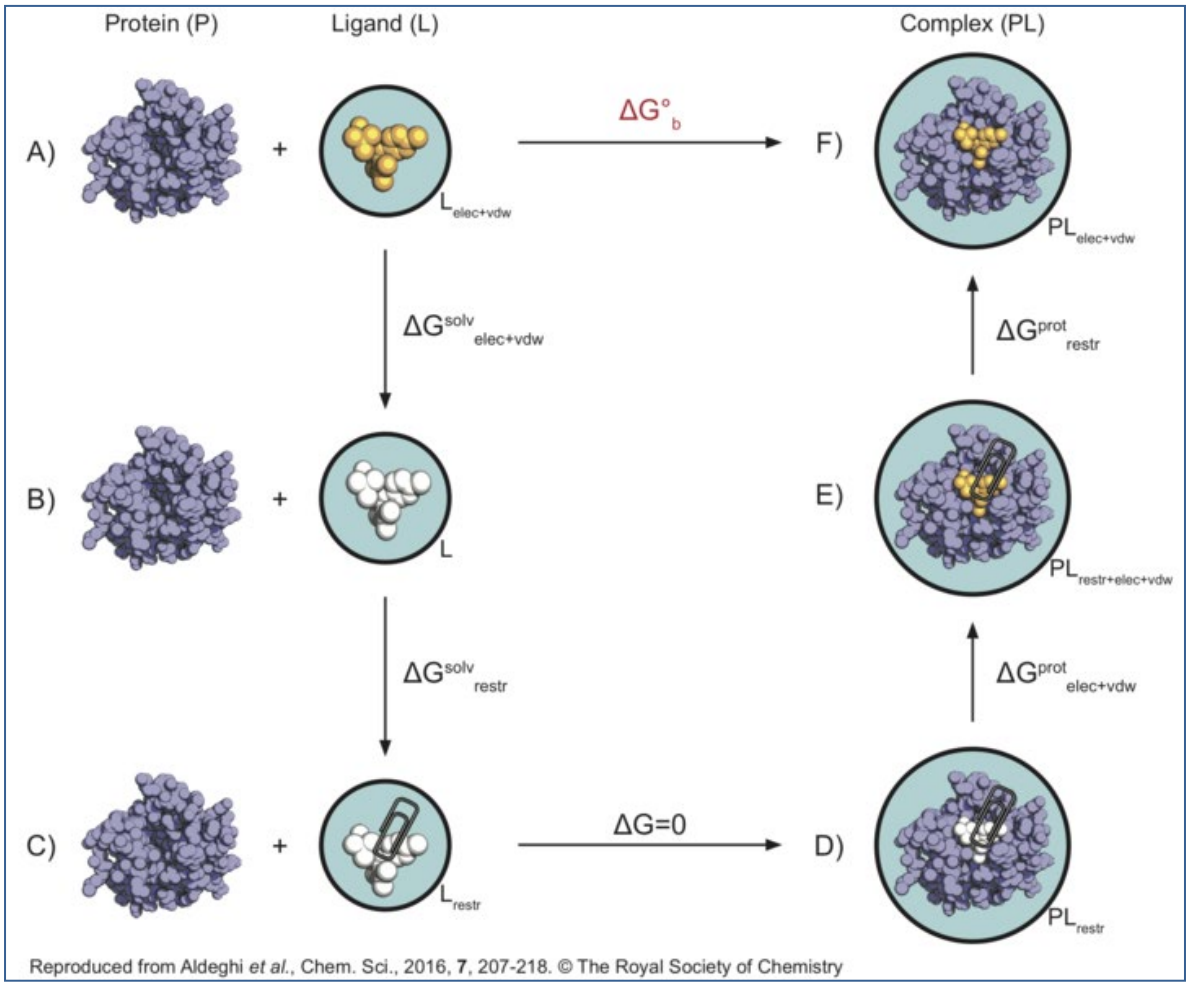

Figure 2: Thermodynamic cycle used to compute absolute binding free energies via alchemical (non-physical) transformations.

The fully interacting ligand (orange) in solution at the top left (A) is transformed into a non-interacting solute (B, white) during a series of equilibrium simulations where its electrostatic and van der Waals interactions are turned off. The ligand is then restrained in space while still non-interacting with the environment (C). This state is equivalent to having the non-interacting ligand restrained within the protein cavity (D). The restrained and non-interacting ligand in complex with the protein has its electrostatic and VdW interactions turned back on (E). The restraints between ligand and protein are finally removed and the final state is the unrestrained and fully interacting ligand in complex with the protein (F). Reproduced from Aldeghi et al.[4]

Although theoretical, exact this procedure requires enhanced sampling to be applicable in the context of large and flexible allosteric protein like pLGICs. Indeed, one needs to sample the conformational ensemble of the receptor in the presence (*holo*) and the absence (*apo*) of the modulatory ligand.

The need for enhanced sampling has been exemplified in several examples, including the case of the T4 Lysozyme where residues in the binding site may adopt multiple and slowly interconverting rotameric states depending on the nature of the ligand. One enhanced sampling method that has been successfully applied to this system is Replica Exchange Solute-Tempering (REST2) where the

ligand and its surrounding, i.e., the protein binding site, are also simulated at higher "temperature" (in fact the protein-ligand interaction energy is scaled down) to enhance the sampling of otherwise slow internal degrees of freedom of the ligand and the protein side chains. So far no out-of-the-box solutions exist to carry on such calculation using open-source tools.

3. Practical implementation

3.1 Software

In the endeavour to provide an open-source and automatised procedure to achieve alchemical binding free-energy calculations with REST2 enhanced-sampling, we chose the Molecular Dynamics engine GROMACS,[5] which implements alchemical free-energy transformations and replica exchange simulations. In addition, GROMACS can be patched with PLUMED,[6] which is a tool for advanced simulation techniques including REST2 to enhance sampling. The protocol has been partially automatised using a set of BASH and Python scripts, in particular the setup of the MD simulations in many replicas, the definition of harmonic restrains to maintain the formation of the complex while uncoupling the ligand, the analysis of the output. This protocol is available on GitHub: https://github.com/AdrienCerdan/FEP-auto).

3.2 Benchmark

Our setup has been benchmarked on a series of synthetic host-guest and protein-ligand complexes for which experimental binding affinities were available, including the challenging T4 Lysozyme (3 complexes), Tankyrase-2 (1 complex), Trypsin (1 complex), and Cucurbit[7]uril (2 complexes).

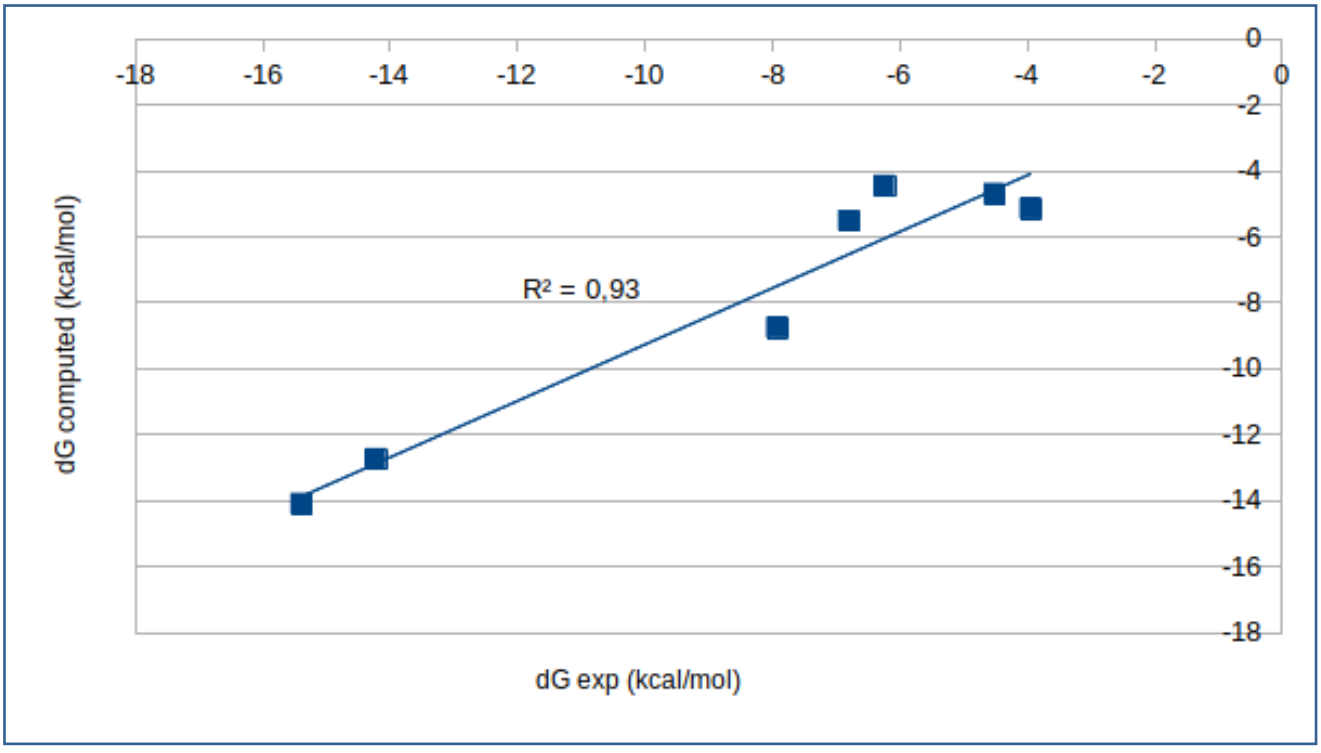

Figure 3: Benchmark of alchemical free energy with GROMACS.

Overall, our protocol can reproduce the experimental binding free energies, including in T4 Lysozyme complexes where enhanced sampling of the rotameric states of some sidechains is required, see Figure 3. In fact, on this test set excellent correlation (R^2=0.93) and mean average error from experiments (MAE=1.15kcal/mol) were obtained, which is in line with the expected accuracy of the methodology.

4. Applications

Before applying the methodology to pLGICs, we targeted the acetylcholine-binding protein (AChBP), which is the soluble analogue of the nAChR. This protein is composed of a pentameric assembly analogue to the extracellular domain of pLGICs, including the binding site of the neurotransmitters. It has the advantages to be smaller and well-characterised by structural biology, as many x-ray structures exist in complexes with several modulatory ligands of the nAChR. Additionally, this protein is thought to be non-allosteric in which case it is anticipated that the differences between the *apo* and *holo* conformations that we need to sample, would be limited to local changes in the binding site, which can be harnessed by the REST2 enhanced sampling technique.

4.1 Initial test on AChBP

We selected 9 complexes of AChBP for which the binding free energy is known experimentally, the X-ray structures were available in the Protein Data Bank, and are pharmacologically relevant in pLGICs, i.e., 3 agonists, 3 partial agonists, and 3 antagonists of the nAChR (PDB code: 2BYQ, 1UW6, 3WIP, 3U8N, 3U8K, 2WNC, 2XYS, 2BYR, 2XNU).

We applied the free-energy simulation protocol described above, which combines Alchemical Free-Energy calculations and REST2 on the 9 selected complexes.

Proceeding with the various complexes we observed the impossibility to converge on the alchemical transformation involving the uncoupling of the ligand from its binding site within the simulation timescale (<10ns per windows). Indeed, in addition to the change in rotameric states of the sidechains we anticipated to tackle with REST2, the binding site also undergoes a motion of the backbone atoms constituting the so-called C-loop, which cap the binding site. The problem here is that in the *holo* state, the C-loop is kept fixed by the presence of the ligand, but in the *apo* state, it freely samples various conformations and degrees of opening. Since it is computationally unfeasible to sample all the conformations of the apo C-loop within the alchemical transformation, we decided to decompose the thermodynamic cycle with two additional steps as presented below.

4.2 Free-energy calculations of the C-loop restrains:

We aimed at computing the free energy of restraining the C-loop conformation in both the *apo* and *holo* states of AChBP and use a restrained C-loop during the alchemical transformation. This procedure is expected to allow faster convergence of the alchemical transformation by reducing the degree-of-freedom to sample, while still being exact from a thermodynamic point of view. For this purpose, we used the Accelerated Weight Histogram (AWH) method,[7] which is an adaptative biasing technique to compute the Potential of Mean Force (PMF) along a given reaction coordinate, here a distance measuring the degree of opening of the C-loop. This method is available in GROMACS and facilitates transition from one step to another within our workflow.

As reaction coordinate, we used the distance between the centre-of-mass of two groups of backbone atoms: (1) the inner-beta sheets of the (-)-subunits and (2) the C-loop of the (+)-subunits. For an efficient sampling, it is recommended to use multiple replicas (or “walkers”) that communicate and share the bias within each other during the simulation. Using 4 and 32 walkers for the *holo* and *apo* state respectively, we were able to obtain converged PMF for confining the C-loop in both *apo* (PDB: 2BYN, see Figure 4) and *holo* (Epibatidine bound to *Aplysia californica* AChBP PDB: 2BYQ, see Figure 5) state of the protein.

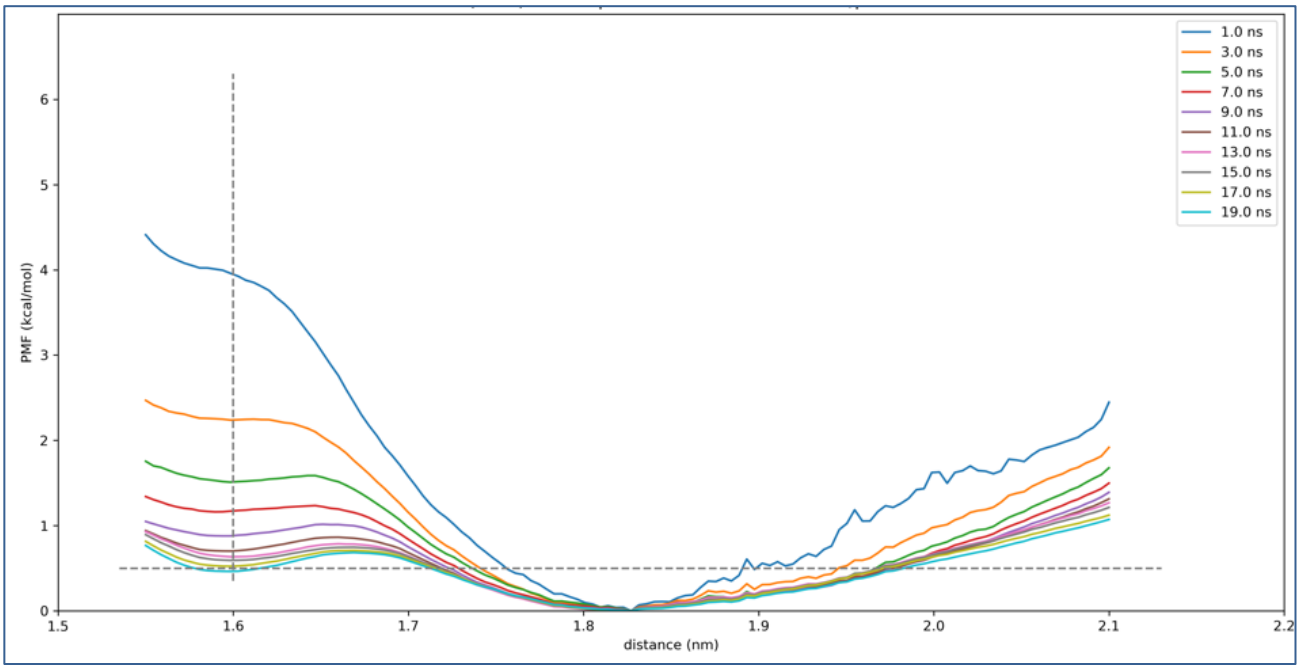

Figure 4: Energetics of opening/closing the C-loop in AChBP *apo*.

The potential of mean force (PMF) calculations by the AWH method show that in the absence of ligands (*apo*) the C-loop prefers to stay open and that closing it to the configuration adopted in the *holo* state of the protein implies a free energy cost of 0.6 kcal/mol. The free-energy calculations converge after 10 ns simulation.

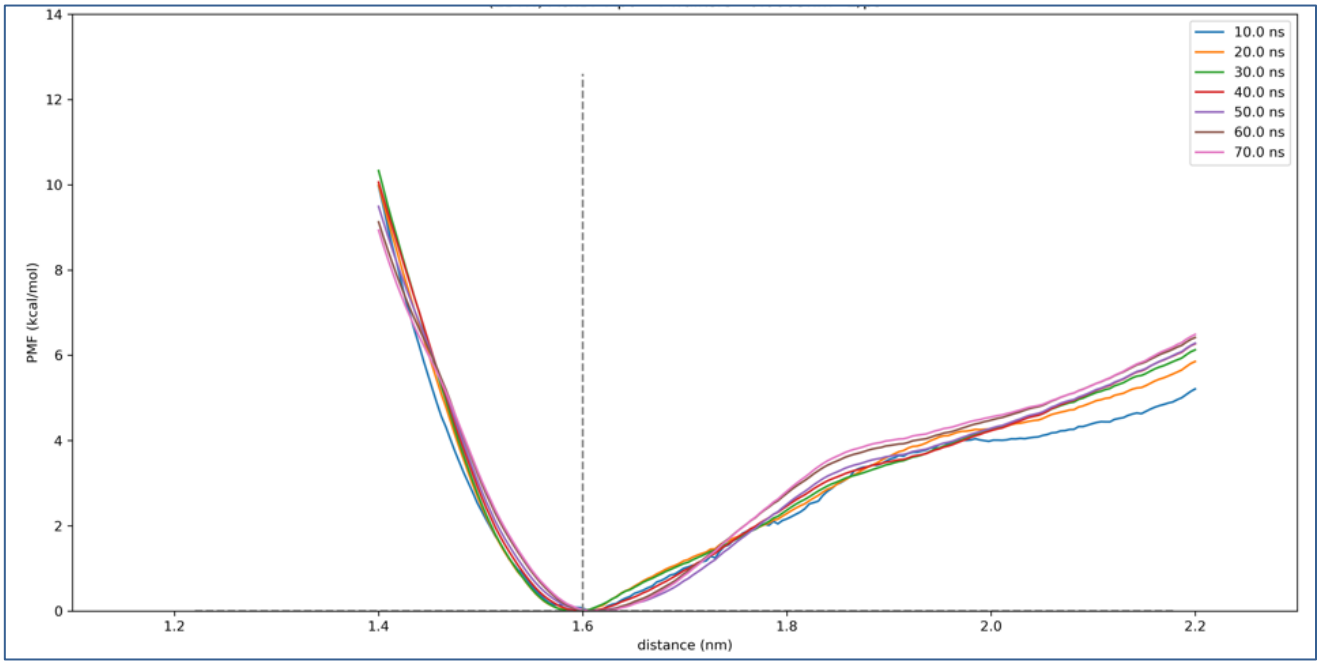

Figure 5: Energetics of opening/closing the C-loop in AChBP *holo*.

The potential of mean force (PMF) calculations by the AWH method show that with epibatidine bound (*holo*) the C-loop is stable in the closed conformation (1.6 nm) The free-energy calculations converge in <10 ns simulation.

As one can see, restraining the C-loop opening to 1.6 Å (i.e., the distance in the X-ray structure of the *holo* state), the free-energy cost is 0 and 0.6 kcal/mol in the *holo* and *apo* states, respectively. With this information we can now proceed to the alchemical transformation to compute the binding free energy in the presence of restrains on the C-loop.

4.3 Binding free-energy calculations in presence of C-loop restrains

Using the extended thermodynamic cycle in Figure 2 along with the introduction of harmonic restraints on the C-loop, we were finally able to obtain converged binding free energy results for the full series of AChBP-ligand complexes.

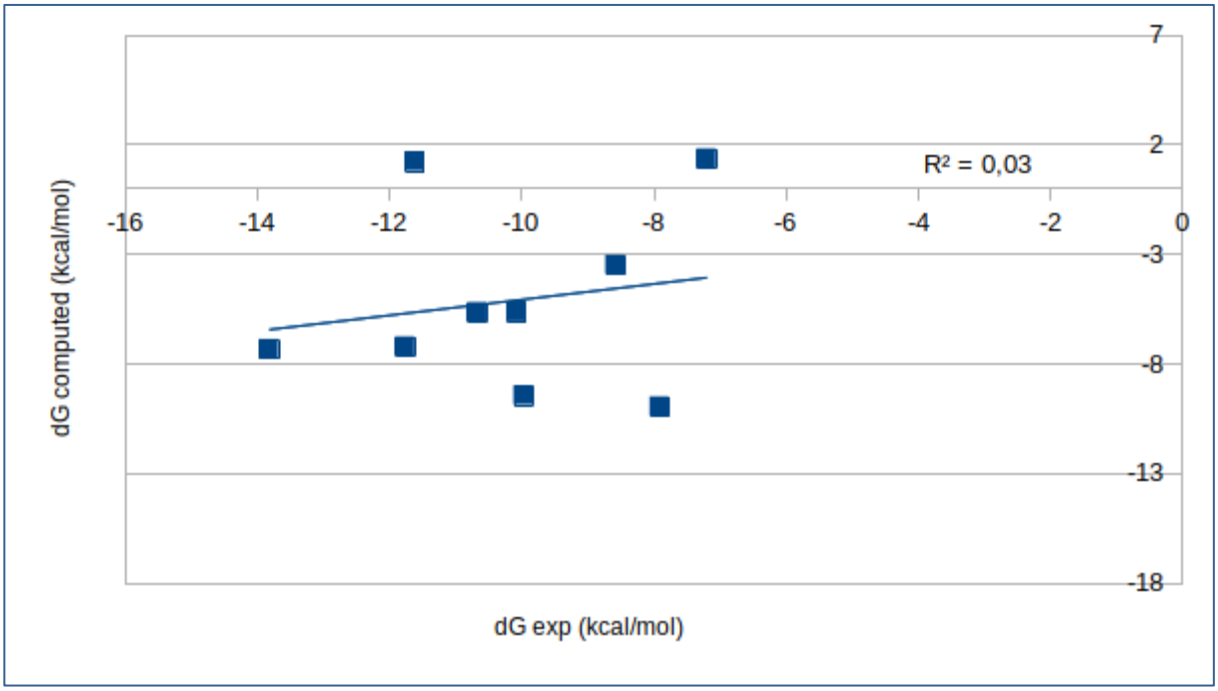

Figure 6: Computed versus experimental binding free energies for a series of 9 AChBP-ligand complexes.

The results are in poor agreement with experiments in regards of both correlation (R^2=0.03) and Mean Average Error (MAE=5.51kcal/mol), see Figure 6. These results are particularly disappointing since the protocol was successful for the benchmark series we investigated before (R^2=0.93 and MAE=1.15kcal/mol).

We hypothesise that force field errors arising from a suboptimal parameterisation of the cation-π interactions, which are critical for protein-ligand recognition in pLGICs, are the source of the failure. Indeed, these interactions are ubiquitous in orthosteric ligands like those investigated here and it was reported that their parameters are unsatisfactory in standard classical force field. To overcome this issue, researchers recently provided updated parameters for a few cation-pi interactions, but the chemical groups we investigated here have not been re-parameterised yet.

5. Conclusions

To advance toward state-based pharmacology in challenging targets like pLGICs we had to setup a complex absolute binding free-energy workflow combining state-of-the-art free-energy simulations such as Alchemical transformations with enhanced sampling techniques like Replica Exchange Solute Tempering (REST) and Accelerated Weight Histogram (AWH). We implemented a partially automatised procedure to carry out these calculations using open-source software, which we provide on GitHub (https://github.com/AdrienCerdan/FEP-auto). These are important steps towards the development of automated strategies for allosteric drug design based on state-based pharmacology based on HPC resources (i.e., ChemFlow).

6. Looking Forward

While our free-energy protocol yielded satisfactory results on benchmark systems, it was unable to provide accurate binding affinities in more challenging protein-ligand systems, like the series of AChBP complexes. Interestingly, our combination of enhanced sampling techniques allows convergence of the calculations, which was not the case otherwise. Our analysis suggests that current modelling of cation-π interactions in classical force fields is suboptimal and should be improved to yield accurate binding free energy results in brain receptors. We anticipate that further development of the force field parameters and extension of the re-parameterisation effort to model these interactions, which are critical for molecular recognition in pLGICs, will be crucial in the future.

The work could not be completed because it was found that absolute binding free energy calculations based on classical MD on the ligand-binding domain of the nicotinic receptor are not accurate enough (i.e. poor correlation with experiments). The initial hypothesis on the design of agonists, antagonists, and allosteric modulators was not confuted. Rather its implementation turned out to be more challenging than expected pointing to shortcomings both in state-of-art ligand-binding free energy calculations or models of protein-ligand interactions for MD (i.e. force fields).

7. References

1. Changeux, J.-P., Amoura, Z., Rey, F. A. & Miyara, M. A nicotinic hypothesis for Covid-19 with preventive and therapeutic implications. *C. R. Biol.* **343**, 33–39 (2020).
2. Melo, G. D. *et al.* Attenuation of clinical and immunological outcomes during SARS-CoV-2 infection by ivermectin. *EMBO Mol. Med.* **13**, (2021).
3. Cecchini, M. & Changeux, J.-P. Nicotinic receptors: From protein allostery to computational neuropharmacology. *Mol. Aspects Med.* **84**, 101044 (2022).
4. Aldeghi, M., Heifetz, A., Bodkin, M. J., Knapp, S. & Biggin, P. C. Accurate calculation of the absolute free energy of binding for drug molecules. *Chem. Sci.* **7**, 207–218 (2016).
5. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
6. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
7. Lindahl, V., Lidmar, J. & Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *J. Chem. Phys.* **141**, 044110 (2014).