

*CDP6 Results SGA2 Year 1 Implementation for accurate HPC-based screening protocols of neuro-drug (D8.6.1 - SGA2)*

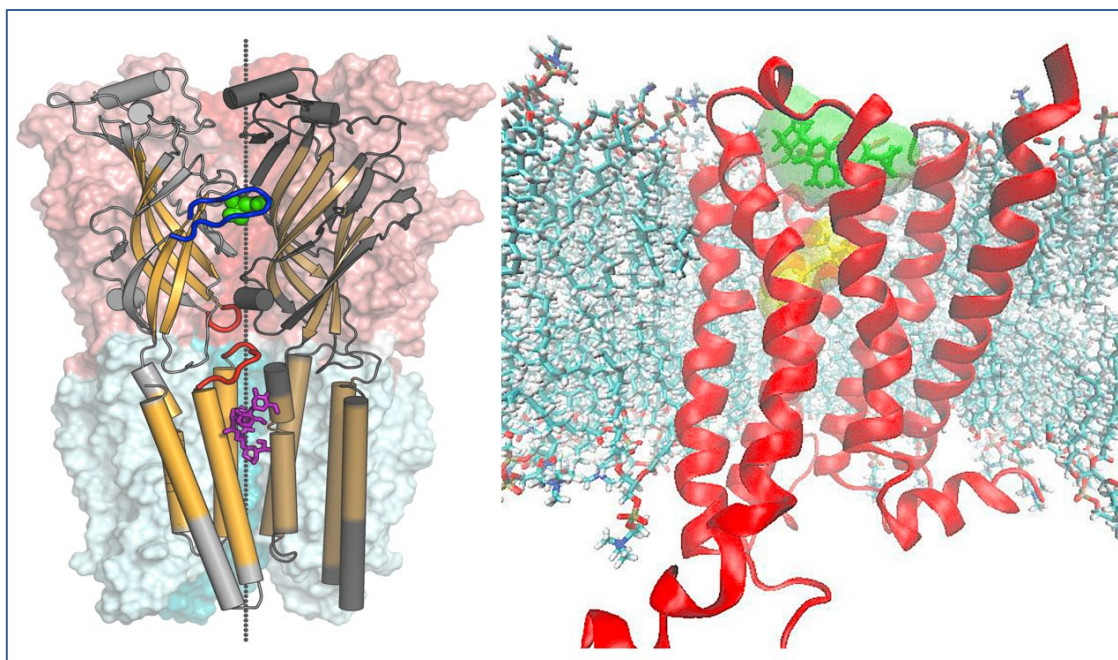


Figure 1: Understanding allosteric regulation in receptors relevant for neuropathologies.

<b>Project Number:</b>	785907	<b>Project Title:</b>	Human Brain Project SGA2
<b>Document Title:</b>	Implementation for accurate HPC-based screening protocols of neuro-drug candidates		
<b>Document Filename:</b>	D8.6.1 (D53.1 D26) SGA2 M12 SUBMITTED 190429.docx		
<b>Deliverable Number:</b>	SGA2 D8.6.1 (D53.1, D26)		
<b>Deliverable Type:</b>	Report		
<b>Work Package(s):</b>	WP8.6		
<b>Dissemination Level:</b>	PU = Public		
<b>Planned Delivery Date:</b>	SGA2 M12 / 31 Mar 2019		
<b>Actual Delivery Date:</b>	SGA2 M13 / 29 Apr 2019; ACCEPTED 23 Jul 2019		
<b>Authors:</b>	Jean-Pierre CHANGEUX, IP (P34), Paolo CARLONI, JUELICH (P20)		
<b>Compiling Editors:</b>	Giulia ROSSETTI, JUELICH (P20), Marco CECCHINI, IP (P34)		
<b>Contributors:</b>	Francesco Luigi GERVASIO, UCL (P82), Chapter contributed: 1, 2, 5, 7 Silvia ACOSTA GUTIÉRREZ, UCL (P82), Chapter contributed: 1, 2, 5, 7 Zoe COURNIA, BRFAA (P120), Chapter contributed: 1, 2, 6, 7		
<b>SciTechCoord Review:</b>	Martin TELEFONT, EPFL (P1)		
<b>Editorial Review:</b>	Annemieke MICHELS, EPFL (P1)		
<b>Description in GA:</b>	<p>CDP6 Results SGA2 Year 1 Implementation for accurate HPC-based screening protocols of neuro-drug candidates.</p> <p>CDP6 Annual compound deliverable Year 1 Overview of Key Results and Impact achieved in M1 to M12, tailored for presentation to the relevant audiences (research/industry/public) References to HBP/SP/CDP objectives and Use Cases for navigation between multiple Sub-Projects Linkage of Results to components/a set of component fact sheets (lower level information: Component ownership, Technology Readiness level, performed Quality Control checks, etc.) This deliverable will also refer to results for SGA2_HPC-based chemical optimisation codes and increasing ligand specificity toward receptors' subclasses</p>		
<b>Abstract:</b>	<p>The aim of this CDP is to develop new strategies for more effective drug treatments of major brain diseases such as Alzheimer's, Schizophrenia, Epilepsy, Parkinson's, glioblastoma and rare diseases using computational models. On the pentameric ligand-gated ion channels (pLGICs), we have characterized the physiological active state of the glycine receptor and completed <math>\mu</math>s-long simulation trajectories for its resting, active, and desensitized state. The identification of 159 compounds from the literature with documented modulatory activity on the GlyR function provided the first complete database (GRALL) for the allosteric modulation of pLGICs. On G-protein-coupled receptors (GPCRs) we have run the firsts virtual screening against human A2A receptor and tested the first set of compounds in <i>in-vitro</i> condition. Positive hits were identified to be further refined. We have further advanced in the understanding of allosteric mechanism of regulations of the human muscarinic acetylcholine M2 receptor. On PI3K<math>\alpha</math>, we have simulated the overactivation of the E545K hotspot mutation, which is implicated in glioblastoma, with enhanced</p>		



	sampling simulations. We have sampled the detachment of the regulatory subunit from the catalytic subunit and shown that the mutant PI3K $\alpha$ E545K can only be found in the overactive state that is implicated in brain cancer. We have also studied the allosteric networks implicated in this transition and identified putative allosteric pockets for protein inhibition.
Keywords:	Molecular Dynamics, drug design, allosteric ligands
Target Users/Readers:	Scientists, Pharma Companies, Biotech Companies, general public

## Table of Contents

<b>1</b>	<b>Overview</b> .....	<b>6</b>
<b>2</b>	<b>Introduction</b> .....	<b>7</b>
<b>3</b>	<b>Key Result KRc6.1 Allosteric modulation of pentameric ligand-gated ion channels</b> .....	<b>8</b>
3.1	Outputs .....	8
3.1.1	Overview of Outputs.....	8
3.1.2	Output 1: Structural and dynamic characterisation of the GlyR physiological states by Molecular Dynamics .....	8
3.1.3	Output 2: The GlyR Allosteric Ligand Library (GRALL) .....	8
3.1.4	Output 3: PhD Thesis (Adrien CERDAN) .....	8
3.2	Validation and Impact.....	9
3.2.1	Actual Use of Output(s) / Exploitation .....	9
3.2.2	Potential Use of Output(s).....	9
3.2.3	Publications .....	9
3.2.4	Measures to Increase Impact of Output(s): Dissemination.....	9
<b>4</b>	<b>Key Result KRc6.2 Allosteric modulation of GPCRs</b> .....	<b>11</b>
4.1	Outputs .....	11
4.1.1	Overview of Outputs.....	11
4.1.2	Output 1: Understanding the mechanism of positive allosteric modulation of GPCR.....	11
4.1.3	Output 2: Screening and Novel clustering workflow on A2A .....	11
4.2	Validation and Impact.....	12
4.2.1	Actual Use of Output(s) .....	12
4.2.2	Potential Use of Output(s).....	12
4.2.3	Publications .....	12
4.2.4	Measures to Increase Impact of Output(s): disseminations.....	12
<b>5</b>	<b>Key Result KRc6.3 Detecting druggable allosteric sites and hidden pockets</b> .....	<b>14</b>
5.1	Outputs .....	14
5.1.1	Overview of Outputs.....	14
5.1.2	Output 1: Physical characterisation of the plasticity and flexibility of two CNS drug-targets adenosine receptor (A2a) and Type-A-gamma-aminobutyric receptor (GABAA) in their native lipid environment.....	14
5.1.3	Output 2: Free energy landscape reconstruction of the A2a receptor activation by means of enhanced sampling simulation .....	14
5.1.4	Output 3: Understanding the influence of the E545K mutation in PI3K $\alpha$ dynamics .....	15
5.1.5	Output 4: Identification of putative allosteric pockets on mutated E545K PI3K $\alpha$ protein ...	15
5.1.6	Output 5: Structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF) .....	15
5.2	Validation and Impact.....	15
5.2.1	Actual Use of Output(s) .....	15
5.2.2	Potential Use of Output(s).....	16
5.2.3	Publications .....	16
5.2.4	Measures to Increase Impact of Output(s): disseminations.....	16
<b>6</b>	<b>Key Result KRc6.4 Targeting the protein-membrane interface for PI3K<math>\alpha</math> allosteric drug discovery</b>	<b>17</b>
6.1	Outputs.....	17
6.1.1	Overview of Outputs.....	17
6.1.2	Output 1: Discovery of the first allosteric PI3K $\alpha$ ligands selective for the PI3K $\alpha$ mutant H1047R. ....	17
6.1.3	Output 2: Defining a set of appropriate Collective Variables (CVs) that best describe the event of activation due to the E545K mutation for parallel tempering Metadynamics simulations.....	18
6.1.4	Output 3: Identify and analyse the full-energy landscape of the WT and mutant PI3K $\alpha$ using multiple walkers simulations. ....	18
6.1.5	Output 4: Identify allosteric pockets that can be exploited for selective E545K inhibition. ....	19
6.2	Validation and Impact.....	19

---

6.2.1	Actual Use of Output(s) .....	19
6.2.2	Potential Use of Output(s).....	19
6.2.3	Publications .....	19
6.2.4	Measures to Increase Impact of Output(s): disseminations.....	20
<b>7</b>	<b>Conclusion and Outlook .....</b>	<b>21</b>



# 1 Overview

Innovative neuropharmacological approaches require a detailed understanding of the molecular and systems-level organisation of the human brain, the causes and mechanisms of diseases, their progression, and the response to treatments. Because of the high level of complexity of the nervous system and of inter-subject variability in molecular brain organisation, behaviour and disease, addressing these issues for any neuropathology appears a daunting task. Indeed, for most neurodegenerative diseases, such as Alzheimer's Disease and Parkinson's Disease, there is currently no cure, in spite of the very large investments from academia and industries.

Classical neuroactive drugs were designed on the basis of their similarity-isosteric competitiveness with compounds of natural origin. The allosteric interaction paradigm, instead, provides the distinction between the orthosteric ligands (binding to the endogenous neurotransmitter sites as agonists or antagonists), and ligands that mediate their effects by interacting with topographically distinct allosteric sites on receptor. Molecular simulations, combined with experimental characterisation, will lead to the discovery of effective allosteric modulators and will help to design new drugs with enhanced selectivity and thus reduced off-target effects.

CDP6 monitors and ensures the delivery of the following objective: Design novel allosteric ligands with possible applications for diagnosis and therapeutic purposes. The WP focuses on three major classes of targets:

- Ligand- and voltage-gated ion channels
- G protein Coupled receptors
- PI3K



## 2 Introduction

CDP6 aims to develop new strategies for more effective drug treatments of major brain diseases, such as Alzheimer's Disease, Epilepsy, Parkinson's Disease, and glioblastoma, using computational models. Traditional neuroactive drugs were designed on basis of their similarity-isosteric competitiveness with physiological receptor substrates. The allosteric interaction paradigm instead, relies on ligands that mediate their effects by interacting at topographically distinct receptor sites from the physiological ones. Molecular simulations, combined with experiments, will lead to the discovery of effective modulators and help to design new drugs with enhanced selectivity and reduced off-target effects.

On the pentameric ligand-gated ion channels (pLGICs): we have characterised the physiological active state of the glycine receptor and completed  $\mu$ s-long simulation trajectories for its resting, active, and desensitised state. The identification of 159 compounds from the literature with documented modulatory activity on the GlyR function provided the first complete database (GRALL) for the allosteric modulation of pLGICs.

On G-protein-coupled receptors (GPCRs): we have run the first virtual screening against human A2A receptor and tested the first set of compounds *in vitro*. Positive hits were identified to be further refined. We have further advanced in the understanding of the allosteric mechanism of regulation of the human muscarinic acetylcholine M2 receptor.

On G-protein-coupled receptors: we have identified the main energetic bottlenecks for activation and some progress has been made towards understanding ligand selectivity. The hetero-pentameric type-A gamma-aminobutyric acid receptor has been physically characterised and collective variables for flux regulation have been identified. In the case of PI3K $\alpha$ , the mechanism of overactivation of the E545K hotspot mutation implicated in glioblastoma has been performed. We have also studied the allosteric networks implicated in this transition and identified putative allosteric pockets for PI3K $\alpha$  inhibition.

New methodologies were developed for drug-design, exploiting newly identified modulatory sites, as well as mechanisms of receptor's activation. The information collected at M12 will be exploited for the identification of chemical scaffolds modulating brain targets, which will be tackled by M24.

## 3 Key Result KRc6.1 Allosteric modulation of pentameric ligand-gated ion channels

Contributors: Marco CECCHINI (SP8 - Unistra/IP) and Jean-Pierre CHANGEUX (SP12 - IP, P34)

### 3.1 Outputs

#### 3.1.1 Overview of Outputs

- 1) Output 1: Structural and dynamic characterisation of the GlyR physiological states by Molecular Dynamics
- 2) Output 2: The GlyR Allosteric Ligand Library (GRALL)
- 3) Output 3: PhD Thesis (Adrien CERDAN)

#### 3.1.2 *Output 1: Structural and dynamic characterisation of the GlyR physiological states by Molecular Dynamics*

Our recent characterisation of the GlyR active state by computational electrophysiology (Cerdan *et al.* Structure 2018) has led to an original structure-function annotation, which provided *bona fide* atomistic models for the resting (R), active (A), and desensitised (D) states of GlyR. Based on this annotation, microsecond long Molecular Dynamics simulations of the three functional states of GlyR embedded in a native lipid bilayer were started and successfully completed. These trajectories contain detailed information on the conformational dynamics of the ligand binding sites during the functional isomerisation of the receptor, which is critically important for developing structural and dynamic models of the conformational transitions and allosteric modulation of ligand-gated ion channels (component 2888). Pocket analysis over time and between states is being carried out to identify the most promising sites for docking studies. The trajectories will be released upon publication and possibly at M24.

#### 3.1.3 *Output 2: The GlyR Allosteric Ligand Library (GRALL)*

Starting from two recent reviews on the allosteric modulation of GlyR function (Lynch 2017, Zeilhofer 2018), we collected chemical structures for 159 compounds with documented modulatory activity on GlyR  $\alpha 1$  and GlyR  $\alpha 3$  into a database named GRALL. The GRALL library contains three agonists, one antagonist, 81 positive allosteric modulators (PAM), 31 negative allosteric modulators (NAM), and 29 decoys. Importantly, for 75 compounds the topographic location of their binding site on the receptor is known from X-ray crystallography studies or site-directed mutagenesis. The availability of a diverse dataset of GlyR modulators is critically important for a quantitative benchmark of new computational strategies for rational allosteric drug design (component 2887). The GRALL library is being prepared and curated (3D structures of all ligands including RESP charges and the receptor in the various physiological states) and will be delivered for public access by M24.

#### 3.1.4 *Output 3: PhD Thesis (Adrien CERDAN)*

Adrien CERDAN defended his PhD dissertation entitled “Exploring synaptic transmission and regulation in ionotropic receptors by molecular dynamics and computational electrophysiology” on 9 February 2019. This thesis was partially funded by HBP.



## 3.2 Validation and Impact

### 3.2.1 Actual Use of Output(s) / Exploitation

### 3.2.2 Potential Use of Output(s)

The coordinates of the GlyR active state isolated by Molecular Dynamics simulations have been published in Structure along with the accompanying paper (see below). The trajectories of the resting and desensitised states will be made available upon publication. These models provide the first complete set of structures for the main physiological states of the same neurotransmitter receptor, which is critically important for a detailed understanding of molecular function in neuronal receptors. In addition, the availability of the first complete dataset of small-molecule modulators of GlyR including decoys (i.e. GRALL) will have a strong impact on the scientific community providing the first quantitative benchmark to test computational strategies for rational design of allosteric drugs. Finally, the establishment of state-based pharmacology approaches, which represents a radical depart from common practices in the Pharma industry, may revolutionize the way drug design is currently conceived not only for neurotransmitter receptors but also for allosteric proteins more generally.

### 3.2.3 Publications

Here is the complete list of publications (disseminations) in SGA2 for this KR.

Cerdan A, Martin N & Cecchini M. “An ion permeable state of the glycine receptor captured by molecular dynamics”, Structure 11, 1555-1562 (2018). <https://doi.org/10.1016/j.str.2018.07.019>

*Significance:* By using an original combination of simulation techniques including computational electrophysiology we provide the closest representation of the physiologically active state of one pLGIC with atomic resolution. These results are crucial for the investigation of allosteric modulation in pLGICs, which is the ultimate goal of CDP6. This work was the subject of a HBP Paper Snapshot in September 2018.

*Output:* 1,3

Wodak SJ, Paci E, Dokholyan N, Berezovsky IN, Horovitz A, Li J, Hilser VJ, Bahar I, Karanicolas J, Stock G, Hamm P, Stote RH, Eberhardt J, Chebaro Y, Dejaegere A, Cecchini M, Changeux JP, Keri D, Barth P, Bolhuis PG, Vreede J, Faccioli P, Orioli S, Ravasio R, Yan L, Brito C, Wyart M, Gkeka P, Rivalta I, Palermo G, McCammon JA, Panecka-Hofman J, Wade RC, Di Pizio A, Niv MY, Nussinov R, Tsai C-J, Jang H & McLeish T. “Allostery in its many disguises: from theory to applications”, Structure 2019, *in press*. <https://doi.org/10.1016/j.str.2019.01.003>

*Significance:* This paper provides an overview of the recent progress and remaining limitations in the understanding of the mechanistic foundations of protein allostery gained from computational and experimental analyses. Also, it demonstrates that we are integral part of the leading community working on protein allostery.

*Output:* 1

### 3.2.4 Measures to Increase Impact of Output(s): Dissemination

- 1) Annual meeting of the Co-Design Project 6 - 30-31 October 2018, Forschungszentrum Juelich (Germany).

*Significance:* This workshop entitled “Modeling for Drug Discovery” attracted 35 international scientists inside and outside CDP6 both from Academia and Industry (MC/GR)



*Output:* This was an overview presentation of the approach used in CDP6 to design novel drugs targeting allosteric sites to modulate the activity of brain receptors.

- 2) BioExcel & European Center of Excellence for Computational Biomolecular Research, 8-9 September 2018, Athens, Greece, keynote lecture JPC

*Significance:* This workshop attracted more than 60 international computer scientists from all over the world concerned by the computational strategies for drug design specifically with Ligand gated ion channels

*Output:* Presentation of the methods and strategies together with the most recent data collected on LGICs.

- 3) Belgian Brain Congress 2018, 18 October 2018, Liege, Belgium, keynote lecture JPC

*Significance:* At this meeting around 300 scientists mostly from Belgium were present from diverse biomedical orientation but interested in basic research

*Output:* A system biology approach for modelling the brain was one of the first public presentations of a new strategy for simulating the brain based upon its organisation in hierarchical levels of organisation with special emphasis on the bottom up importance of the molecular level.

## 4 Key Result KRc6.2 Allosteric modulation of GPCRs

Contributors: Giulia ROSSETTI (SP2 - JUELICH, P20) and Paolo CARLONI (SP6 - JUELICH, P20)

### 4.1 Outputs

#### 4.1.1 Overview of Outputs

- 1) Output 1: Understanding the mechanism of positive allosteric modulation of GPCR
- 2) Output 2: Screening and Novel clustering workflow on A2A

#### 4.1.2 Output 1: Understanding the mechanism of positive allosteric modulation of GPCR

We clarified the molecular basis of positive allosteric modulation and identified the key elements, which make this phenomenon possible. We used as a test case the M2 muscarinic acetylcholine receptor (mAChR), which was shown to have an allosteric vestibular site and whose dynamic properties were extensively studied by us, as well as by others. Three different crystallised complexes of this receptor were examined and simulated by Molecular Dynamics: the orthosteric antagonist-bound complex, the orthosteric agonist-bound complex, and the positive allosteric modulator-bound complex in the presence of the same orthosteric ligand. Lead by the idea of allosteric modulation as communication means among different regions of a macromolecule, we have modelled this as a vibrational energy exchange among protein residues. Residues responsible for the energy flux increase have been identified and, as expected, they seem to play fundamental roles for either allosteric or orthosteric binding. The last finding provides valuable information for further allosteric drug development (Maggi *et al.* 2018) (Maggi *et al.* 2019 - in preparation).

Component Field	Data
ID	2890
Title	<i>In silico</i> screening of ligands binding to the receptor
Type	report
Leader	Giulia ROSSETTI, Paolo CARLONI
(Release) Name	Maggi <i>et al.</i> 2018
Release URL	<a href="http://doi.org/10.1021/acs.jpcllett.8b02380">http://doi.org/10.1021/acs.jpcllett.8b02380</a>
Effective (Release) Date	14 October 2018

#### 4.1.3 Output 2: Screening and Novel clustering workflow on A2A

In detail, we are building an in-house platform to model GPCRs, simulate those at multi-scale level, screen new molecules and cluster those accordingly to their similarity and novelty.

The main goal of CDP6 SGA2 M24 Deliverable D8.6.2 is to develop of new HPC approaches for the rational optimisation of drugs.

We have developed a combined HPC and machine-learning protocol for structure-activity model building. As a test case we used the Adenosine receptor A2A.



Component Field	Data
ID	2889
Title	<i>In silico</i> screening of selective Ligands against human Adenosine Receptor Subtypes.
Type	report
Leader	Giulia ROSSETTI, Paolo CARLONI
(Release) Name	Cao <i>et al.</i> 2018
Release URL	<a href="http://doi.org/10.3390/molecules23102616">http://doi.org/10.3390/molecules23102616</a>
Effective (Release) Date	12 October 2018

## 4.2 Validation and Impact

### 4.2.1 Actual Use of Output(s)

Output 1 and 2 are currently being used to develop novel regulators of receptor activities through virtual screening campaigns validated by experiments.

### 4.2.2 Potential Use of Output(s)

The results obtained in Output 1 and 2 will provide novel tracers for these receptors with improved specificity, as well as they will open the way to targeted therapeutic strategies

### 4.2.3 Publications

Maggi, L., Carloni, P., & Rossetti, G. (2018). Vibrational Energy in Proteins Correlates with Topology. *The Journal of Physical Chemistry Letters*, 9(22), 6393-6398. <http://doi.org/10.1021/acs.jpcllett.8b02380>

*Significance:* It establishes for the first time a link between vibrational energy and topology in proteins

*Output:* Novel methods to quantify exchange of vibrational energy that is relevant in protein function.

Cao, R., Giorgetti, A., Bauer, A., Neumaier, B., Rossetti, G., & Carloni, P. (2018). Role of Extracellular Loops and Membrane Lipids for Ligand Recognition in the Neuronal Adenosine Receptor Type 2A: An Enhanced Sampling Simulation Study. *Molecules*, 23(10), 2616. <http://doi.org/10.3390/molecules23102616>

*Significance:* It establishes the membrane-dependent allosteric modulation on GPCR activation

*Output:* characterisation of membrane role in the ligand binding process.

### 4.2.4 Measures to Increase Impact of Output(s): disseminations

1) Annual meeting of the Co-Design Project 6 (CDP6) at Forschungszentrum Juelich (Germany).

*Significance:* This workshop entitled “Modeling for Drug Discovery” attracted 35 international scientists both from Academia and Industry (MC/GR).

*Output:* Network of collaboration for the future SGA3 was established.



- 2) KeyLab International Workshop on "Recent computational and experimental advances in molecular medicine", 27-29 June 2018: Ho Chi Minh City, Vietnam  
*Significance:* This workshop brought HBP ideas and results also outside Europe  
*Output:* Student and collaborators recruited.
- 3) CECAM workshop "Physiological role of ions in the brain: towards a comprehensive view by molecular simulation", 21 May 2018: Pisa, Italy  
*Significance:* This workshop for crucial to attract expert in the role of physiological ions in the the brain  
*Output:* Students and collaboration recruited.

## 5 Key Result KRc6.3 Detecting druggable allosteric sites and hidden pockets

Contributor: Francesco Luigi GERVASIO (SP8 - UCL, P82)

### 5.1 Outputs

#### 5.1.1 Overview of Outputs

- 1) Output 1: Physical characterisation of the plasticity and flexibility of two CNS drug-targets adenosine receptor (A<sub>2a</sub>) and Type-A-gamma-aminobutyric receptor (GABAA) in their native lipid environment
- 2) Output 2: Free energy landscape reconstruction of the A<sub>2a</sub> receptor activation by means of enhanced sampling simulation
- 3) Output 3: Understanding the influence of the E545K mutation in PI3K $\alpha$  dynamics
- 4) Output 4: Free energy landscape differences of active and inactive states induced by mutation E545K in PI3K $\alpha$
- 5) Output 5: Structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF).

This output refers to component C3002.

#### 5.1.2 *Output 1: Physical characterisation of the plasticity and flexibility of two CNS drug-targets adenosine receptor (A<sub>2a</sub>) and Type-A-gamma-aminobutyric receptor (GABAA) in their native lipid environment*

Starting from a conformational ensemble of crystal structures of the adenosine receptor (A<sub>2a</sub>) in different states: active, inactive and intermediates, we performed microsecond-long molecular dynamics simulations to physically characterise the plasticity and flexibility of the receptor, as well as to identify the energetic bottlenecks for interconversion between states. We used state-of-the-art force fields, namely amber-disp with tip4p-d water.

For the type-A-gamma-aminobutyric receptor (GABA<sub>A</sub>), we performed microsecond-long molecular dynamics simulations (with state-of-the-art force fields) starting from the a hetero-pentameric structure of the receptor embedded in a realistic synaptic membrane. We characterised important physical properties for nanopores conductance such as flexibility and internal electrostatics: both the internal electrostatic potential and the internal electric field; as well as the region of potential decay inside the receptor.

#### 5.1.3 *Output 2: Free energy landscape reconstruction of the A<sub>2a</sub> receptor activation by means of enhanced sampling simulation*

To calculate the relative populations and energy barriers separating the conformational repertoire of the A<sub>2a</sub> receptor, we derived an activation collective variable from the molecular dynamics simulations, and reconstructed the free energy landscape for the receptor activation from a 16 replicas well-tempered-ensemble, parallel-tempering Metadynamics simulations.

### **5.1.4 Output 3: Understanding the influence of the E545K mutation in PI3Ka dynamics**

To better understand the differences in the dynamics, which arise from the introduction of the E545K mutation, microsecond long unbiased Molecular Dynamics simulations of the WT and mutant protein have been carried out and extensive analysis on their different biophysical properties has been completed.

### **5.1.5 Output 4: Identification of putative allosteric pockets on mutated E545K PI3Ka protein**

The aim of this part of the HBP is to fully describe the mechanism of E545K mutant overactivation by studying the energy landscape of this oncogenic mutant and compare it with the WT protein. For the efficient sampling of the energy landscape of the WT and mutant PI3Ka, parallel-tempering Metadynamics simulations with 20 replicas at increasing temperatures (300 to 319 K, spaced every 1 K) for both systems have been performed. With this, we reconstructed the conformational free energy landscape of both the WT and mutant protein and calculate the relative energy and population of the active and mutation-induced inactive states. Metastable states in the free energy landscape were analysed to identify putative allosteric binding pockets distinctive for the mutant protein only, which do not exist in the WT.

### **5.1.6 Output 5: Structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF)**

The Nerve growth factor (NGF) is an important neurotrophic factor involved in the regulation of cell differentiation and survival of target neurons. Expressed as a pro-NGF precursor, NGF is matured by furin-mediated protease cleavage. Increasing evidence suggests that NGF and pro-NGF have distinct functional roles, but while the structure of mature NGF is available, little was known about that of the pro-domain because of its partially unfolded status. We exploited an ad hoc hybrid strategy based on nuclear magnetic resonance and MD simulations validated by small-angle X-ray scattering. This will allow us to gain novel insights on the prodomain, both in isolation and in the context of proNGF.

## **5.2 Validation and Impact**

### **5.2.1 Actual Use of Output(s)**

Molecular dynamics simulations of G-protein coupled receptors systems are important to understand their conformational plasticity. Notwithstanding recent breakthroughs in crystallography and the increasing number of GPCR structures, the mechanism of their selective activation is unclear. GPCR signalling involves allosteric effects, significant conformational changes and the recruitment of specific intracellular partners. Thus, static crystal structures need to be complemented by other techniques, such as simulations, to fully understand their activation mechanisms. The reconstructed free energy landscapes for the activation of the receptor are useful to understand their conformational repertoire as the on/off paradigm seems to be not representative of the real biophysics of the system.

In the case of the type-A gamma-aminobutyric receptor, there are no available data about the dynamics of the hetero-pentameric system, as its crystal structure has only been released this year. Understanding the dynamics and conformational plasticity of the receptor, as well as its intrinsic

physical characteristic like electrostatics and conductivity from and all atom point of view, will be important to design safer drugs targeting the flux regulation of the channel.

Studying the free energy of the PI3K $\alpha$  E545K overactivation mutant landscape and comparing it to the WT protein will lead to the identification of cryptic, allosteric pockets that can be used for selective inhibition of the mutant protein.

The simulations of pro-NGF have been validated by a range of different structural and biophysical methods (NMR, SAXS) and might eventually lead to new strategies to modulate NGF activity.

### **5.2.2 Potential Use of Output(s)**

The free energy landscape for the activation of G-protein coupled receptor could potential guide the design of new biased ligands capable to selectively activate the recruitment of the G-protein or arresting, increasing the safety of the designed drugs.

The molecular dynamics of the hetero-pentameric type-A gamma-aminobutyric acid receptor together with the physical observables derived from the dynamics, will help to identify the energetic bottlenecks and key conformational changes among the close-desensitised-open states of the receptor. Moreover, as the receptor dynamics have been simulated in a realistic synaptic environment, key interactions with the environment might be targeted pharmacologically.

The identified cryptic allosteric pockets in the E545K PI3K $\alpha$  mutant will be used to design selective inhibitors.

### **5.2.3 Publications**

### **5.2.4 Measures to Increase Impact of Output(s): disseminations**

The outputs will be (and have been) presented in international conferences targeting a wide audience.

FLG has delivered two invited talks at the American Chemical Society Spring and Fall meeting in 2019 presenting the results obtained across the different systems.



## 6 Key Result KRc6.4 Targeting the protein-membrane interface for PI3K $\alpha$ allosteric drug discovery

Contributor: Zoe COURNIA (SP8 - BRFAA, P120)

Within KRc6.4 a rational, structure-based methodology will be designed, which will identify small molecules that modulate protein-membrane interactions, following the identification of druggable allosteric pockets on the protein-membrane interface. This concept will be applied on identifying novel PI3K $\alpha$  allosteric inhibitors against mutated PI3K $\alpha$ , implicated in glioblastoma. (C3003)

### 6.1 Outputs

#### 6.1.1 Overview of Outputs

- 1) Output 1: Discovery of the first allosteric PI3K $\alpha$  ligands selective for the PI3K $\alpha$  mutant H1047R.
- 2) Output 2: Defining a set of appropriate Collective Variables (CVs) that best describe the event of activation due to the E545K mutation for parallel tempering Metadynamics simulations
- 3) Output 3: Identify and analyse the full-energy landscape of the WT and mutant PI3K $\alpha$  using multiple walkers simulations
- 4) Output 4: Identify allosteric pockets that can be exploited for selective E545K inhibition

#### 6.1.2 Output 1: Discovery of the first allosteric PI3K $\alpha$ ligands selective for the PI3K $\alpha$ mutant H1047R.

Molecular Dynamics simulations were used to identify the most probably conformations of the wild-type (WT) PI3K $\alpha$  and the H1047R PI3K $\alpha$  mutant, which is a hotspot mutation in glioblastoma. Putative allosteric pockets were predicted on these conformations using available software in the mutant and WT proteins. The pockets that were identified only on the mutant structure were tested for correlating their motion with the active site in order to verify whether they are implicated in allosteric motions. On these pockets, computer-aided drug design was performed using docking. Careful post-processing of the docked poses was ensued by filtering top-scoring compounds not only for high docking score but also for predicted solubility, cell permeability, number of metabolites, viability of docked pose by means of steric hindrance, and clustering to ensure maximal chemical diversity.

Selected compounds were purchased and tested using an *in vitro* assay containing PI3K $\alpha$  wild-type or PI3K $\alpha$  H1047R mutant and cancer liposomes containing PIP<sub>2</sub>, and PI3K $\alpha$  inhibition was measured. Several selective inhibitors of H1047R PI3K $\alpha$  compared to the wild-type were discovered. One of the compounds, PI3K-010 showed an 11-fold specificity for the H1047R PI3K $\alpha$  mutant. PI3K-010 was then tested for PI3K $\alpha$  inhibition in presence of high (2mM) and low (100 $\mu$ M) ATP concentrations, to verify whether the inhibitor is competitive or non-competitive. The measured IC<sub>50</sub> of PI3K-010 remained unchanged regardless of high or low ATP concentration, leading to the hypothesis that this inhibitor is non-competitive.

PI3K-010 was then tested for modulation of the PI3K $\alpha$ -membrane interaction using Surface Plasmon Resonance, where liposomes from cancer patients were attached on the SPR chip and PI3K $\alpha$  WT or H1047R mutant were preincubated with PI3K-010 and inserted afterwards in the SPR chip. The SPR experiment showed a dramatic increase of the PI3K $\alpha$  protein on the liposomes in the presence of PI3K-010 compared to other competitive PI3K $\alpha$  inhibitors such as Wortmannin.

The *in vivo* activity of PI3K-010 was tested in xenografts, where cells from a cancer cell line are implanted into immunodeficient mice. PI3K-010 was also tested in genetically modified mice, engineered to harbour the mutant H1047R PI3K $\alpha$ , which develop cancer. Using PET/CT imaging, images acquired at the beginning and end of the two-week cancer treatment with the PI3K-010 compound, tumour necrosis and reduced proliferation was observed in the mutants compared to the controls (proliferation index 15.0 $\pm$ 3.3% and 69.9 $\pm$ 5.1%, respectively). Compound PI3K-010 was optimised to reach > 100-fold selectivity. These results were submitted as a patent application.

These studies are linked to C3003.

### **6.1.3 Output 2: Defining a set of appropriate Collective Variables (CVs) that best describe the event of activation due to the E545K mutation for parallel tempering Metadynamics simulations**

The canonical distribution of a system sampled via Molecular Dynamics (MD) carries full information about its thermodynamic properties. However, this distribution is of very little use, as the space in which it is defined is huge. For this reason, molecular systems are often analysed in terms of collective variables (CVs) rather than atomic coordinates. A CV is a function of the atomic coordinates that is capable of describing the physics behind the process under investigation. An appropriate CV or an appropriate set of CVs should be able to distinguish between the initial, final, and intermediate states clearly, should be able to describe all the slow events that are relevant to the process of interest, and their number should not be too large, otherwise it will take a very long time to fill the free energy surface. Using the results from the unbiased simulations for PI3K $\alpha$  E545K with the AMBER99SB-ILDN, we tested several different combinations of collective variables, until we defined a set of CVs that could then be used to bias the system using metadynamics (metaD) simulations. Specifically, we used Steered Molecular Dynamics (SMD) and Well Tempered Metadynamics (WT-metaD) simulations to define a set of appropriate CVs, which best describe the event of activation due to the E545K mutation.

By performing these simulations, we were able to identify that the optimal CVs are two: CV1, which describes the distance between the centres of mass of the nSH2 and helical domains, and CV6, which describes the distance in contact map space to the “open” (detached) state. This study enabled further the parallel tempering metaD (PT-metaD) simulations of the WT that were then carried out biasing both CVs and are described below.

In order to efficiently study the mechanism of E545K mutant overactivation, parallel-tempering Metadynamics simulations with 20 replicas at increasing temperatures (300 to 319 K, spaced every 1 K) for both systems have been performed using the choice of Collective Variables (CVs) that were defined as described above. Using this methodology, the free energy landscape of the WT and E545K PI3K $\alpha$  protein were investigated and the relative energy and population of the active and mutation-induced inactive states were calculated. (C3003)

### **6.1.4 Output 3: Identify and analyse the full-energy landscape of the WT and mutant PI3K $\alpha$ using multiple walkers simulations.**

The studies of parallel-tempering Metadynamics simulations indicated that the mutant E545K PI3K $\alpha$  has only one stable free energy minimum, which is the open state, contributing to overactivation of the kinase. However, due to the complexity of the system and the free-energy landscape the calculation did not converge, although the qualitative picture was clear. For this reason, multiple simulations (walkers) of the same system were run in parallel using metadynamics on the same set of collective variables. In this simulation, the deposited bias is shared among the replicas in such a way that the history-dependent potential depends on the whole history. This calculation showed as

well that the E545K mutant possesses only one stable free energy minimum in the open (overactive) state in contrast to the wild-type simulation. The wild-type simulation of PI3K $\alpha$  showed that the protein in its normal state explores two states: the inactive state and the active state as known from experimental results. This simulation confirmed the result of the parallel-tempering Metadynamics simulations above. (C3003)

### **6.1.5 Output 4: Identify allosteric pockets that can be exploited for selective E545K inhibition.**

To understand the role of residue E545K in PI3K $\alpha$  allosteric regulation, we have constructed network models obtained from our MD simulations. For our analyses, we used the Dynamical Network Analysis method. Based on this model, allosteric signals are dependent on positional correlations of protein residues, and correlated motion is used to generate a weight of the signal transmitted through two residues. Apart from identifying allosteric pathways, the method also applies hierarchical clustering using the Girvan-Newman algorithm to cluster residues whose motion is highly correlated in so-called “communities” that are highly intra-connected but loosely inter-connected. The communication between different communities passes through specific residue interactions that form critical edges. A detailed analysis of the inter-domain communication network has been performed for the WT (closed-state) and mutant (open-state).

Next, metastable states in the free-energy landscape were analysed to identify putative allosteric binding pockets distinctive for the mutant protein only, which do not exist in the WT. These calculations were performed with SiteMap (Schrodinger) that mapped and evaluated the sites based on the concave, hydrogen bond donors, acceptors and hydrophobic characteristics. These pockets were subsequently tested using normal mode analyses to identify whether they are implicated in large scale motions of functional domains of the protein and/or the inter-domain communication network that was identified through the allosteric networks above. The next step of the work will be to identify new E545K-specific inhibitors of PI3K $\alpha$ . (C3002 and C3003)

## **6.2 Validation and Impact**

### **6.2.1 Actual Use of Output(s)**

Output 1 successfully developed novel small molecule inhibitors of the mutant H1047R form of PI3K $\alpha$  which is involved in glioblastoma. These inhibitors are the first allosteric regulators reported for PI3K $\alpha$ . Moreover, the modulation of the PI3K $\alpha$ -membrane interface by PI3K-010 was shown using SPR.

Output 2-4 are currently being used to develop novel regulators of the other hotspot PI3K $\alpha$  mutation, E545K.

### **6.2.2 Potential Use of Output(s)**

The results obtained in Output 1, 2, 3, 4 have the potential to translate targeted therapeutics.

### **6.2.3 Publications**

Z. Cournia, S Christoforidis, A Kapela, E Couladouros, A Efstratiadis, “Method of preparation and use of phosphoinositide 3-kinase inhibitors in treating cancer”. OBI Patent Application number 20180100392.

*Significance:* The invention discloses the first ever allosteric inhibitors of PI3K $\alpha$ , which are selective for the mutant H1047R form, which is prevalent in glioblastoma. The patent is not public for 18



months. Once it becomes public we will be able to measure its impact. The potential impact would be enormous in case the compounds progress into clinical trials for glioblastoma.

*Output:* The invention relates to novel compounds that are useful in medicine, specifically in treating or preventing cancerous diseases in a mammal, preferably in humans and to pharmaceutical compositions comprising such compounds, optionally together with other pharmaceutically active compounds, as well as pharmaceutical formulations comprising such compounds or pharmaceutical compositions. These candidate drugs are an excellent example of applying a personalised medicine pipeline for diseases of the brain, whereby patients harbouring the H1047R mutation in glioblastomas will benefit. The compounds have the potential to progress to the clinic and translate bench results to bedside.

## 6.2.4 Measures to Increase Impact of Output(s): disseminations

Some /the main three dissemination activities of this KR:

HBP Conference Presentations (Zoe COURNIA)

- 1) GCC 2018, 14th German Conference on Chemoinformatics, “Using Free Energy Perturbation Calculations to Predict Relative Binding Affinities in Drug Design”. 11-13 November 2018, Mainz, Germany, keynote speaker.
- 2) CECAM Workshop on Multiscale simulations of allosteric regulatory mechanisms in cancer-associated proteins and signaling protein networks, “Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations”, 15-17 October 2018, Lugano, Switzerland, invited talk.
- 3) 22nd EuroQSAR Conference “Translational and Health Informatics”, “Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations”, 16-20 September 2018, Thessaloniki, Greece, invited talk.



## 7 Conclusion and Outlook

The knowledge of allosteric transitions of neurotransmitter receptors and ion channels at the atomistic level has a profound impact on our understanding of the operations carried out by the billions of nerve cells of our brain, reacting to chemical signals that mediate information processing in the brain from the molecular to the cognitive level. Therefore, for brain modelling, it is important to start at the molecular level, a level that is often under-evaluated or absent from artificial intelligence/neuromorphic modelling. It offers another view on neuronal plasticity, besides offering a broad range of new opportunities for drug design. Furthermore, the sensitivity and responsiveness of a given brain disease to a spectrum of drugs is a critical feature of disease classification and their so-called ontology.

CDP6 therefore covers an important aspect of future medical informatics research: the establishment of causal links between human genome data and the origins of brain diseases. The ultimate goal of CDP6 is to develop new strategies for more effective drug treatments of major brain diseases such as Alzheimer's Disease, Schizophrenia, Epilepsy, Parkinson's Disease, glioblastoma and rare diseases, by using molecular simulations and HPC resources.

At the current stage, several CNS drug targets (i.e. two pentameric-ligand gated ion channels, two GPCRs, and two protein brain cancer oncogenes) have been successfully explored by Molecular Dynamics. The microsecond-long trajectories are currently analysed to identify and characterise putative modulatory sites in these brain receptors, which will be soon targeted in virtual screening campaigns. In addition, new methodologies to develop original strategies for allosteric drug design, identify modulatory sites at the protein-membrane interface, and explore receptor's activation with atomic resolution are being developed. The information collected at M12 along with the developed methodologies are now ready to be exploited for the identification of chemical scaffolds for the allosteric modulation of very important brain targets, which will be tackled by M24.