

HPC-based chemical optimisation codes increasing ligand specificity toward receptors' subclasses
(D8.6.2 - SGA2)

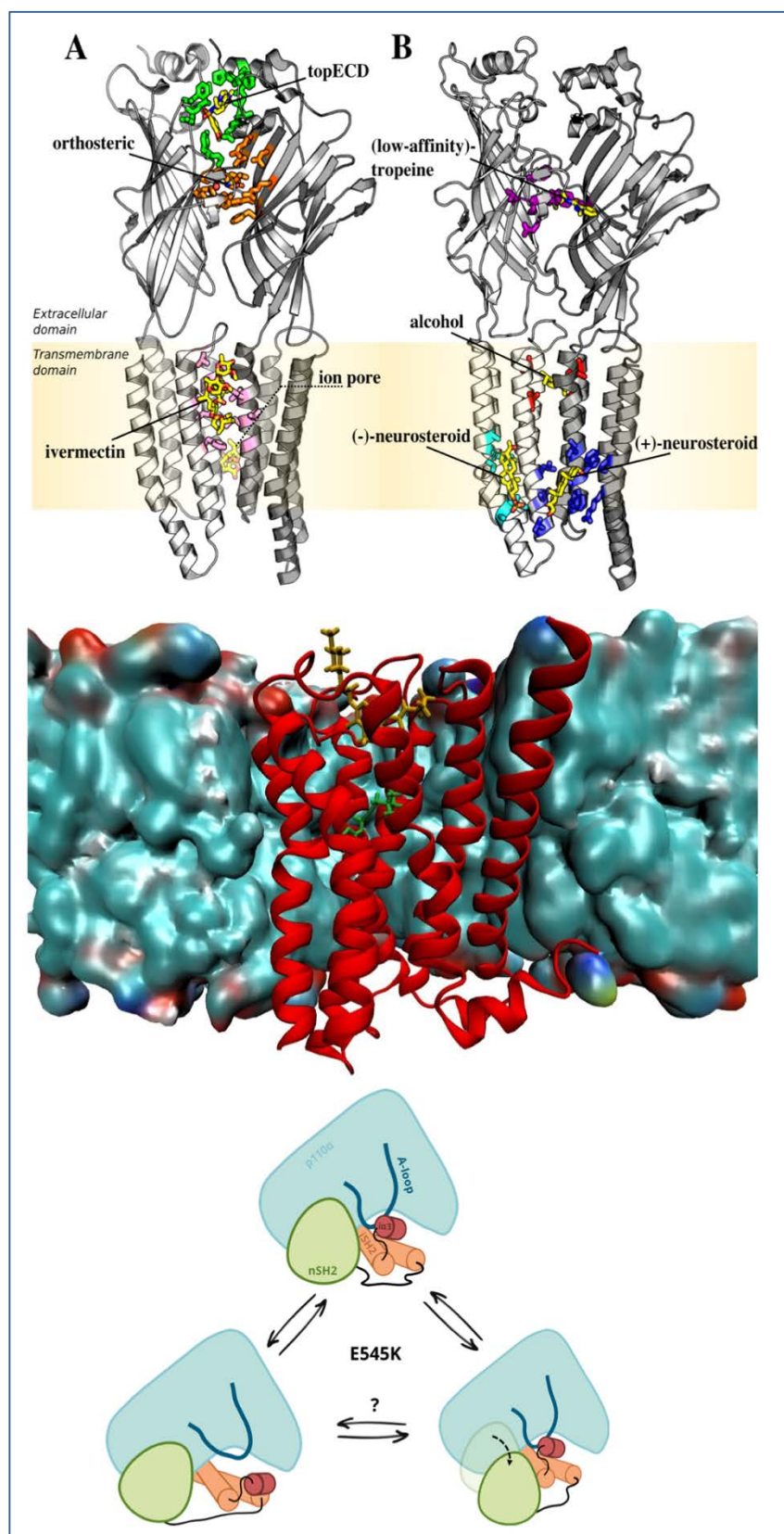


Figure 1: Allosteric regulation in therapeutic targets relevant for neuropathologies

Project Number:	785907	Project Title:	Human Brain Project SGA2
Document Title:	CDP6 Results SGA2 Year 2 HPC-based chemical optimisation codes increasing ligand specificity toward receptors' subclasses		
Document Filename:	D8.6.2 (D53.2 D45) SGA2 M24 ACCEPTED 201006.docx		
Deliverable Number:	SGA2 D8.6.2 (D53.2, D45)		
Deliverable Type:	Report		
Work Packages:	WP8.6		
Key Result(s):	KRc6.1, KRc6.2, KRc6.3, KRc6.4		
Dissemination Level:	PU = Public		
Planned Delivery Date:	SGA2 M24 / 31 Mar 2020		
Actual Delivery Date:	SGA2 M26 / 26 May 2020; approved 31 Jul 2020; resubmitted 2 Oct 2020; accepted 6 Oct 2020		
Author(s):	Giulia ROSSETTI, JUELICH (P20); Marco CECCHINI, IP(P34); Francesco Luigi GERVASIO, UCL (P82), Zoe COURNIA, BRFAA (P120); Paolo CARLONI, JUELICH (P20), Jean-Pierre CHANGEUX, IP (P34)		
Compiled by:	Giulia ROSSETTI, JUELICH (P20); Marco CECCHINI, IP(P34)		
Contributor(s):	Giulia ROSSETTI, JUELICH (P20), Paolo CARLONI, JUELICH (P20), Section 1, 2, 4, 7 Marco CECCHINI, IP (P34), Jean-Pierre CHANGEUX, IP (P34), Section 1, 2, 3, 7 Francesco Luigi GERVASIO, UCL (P82), Section 1, 2, 5, 7 Zoe COURNIA, BRFAA (P120), Section 1, 2, 6, 7		
SciTechCoord Review:	Martin TELEFONT, EPFL (P1)		
Editorial Review:	Annemieke MICHELS, EPFL (P1)		
Description in GA:	For consistent presentation of HBP results, SGA2 M24 Deliverables describing the accomplishments of an entire SP, WP or CDP have been prepared according to a standard template, which focuses on Key Results and the outputs that contribute to them. Project management elements such as Milestones and Risks will be covered, as per normal practice, in the SGA2 Project Periodic Report.		
Abstract:	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. In KRc6.1, on the Glycine receptor, target for chronic inflammatory pain, we have initiated the first virtual high-throughput screening campaign targeting a modulatory site for potentiation. In KRc6.2 we have initiated the first virtual high-throughput screening against selected classes of GPCRs, relevant for neuropathologies, implementing kinetics and thermodynamics information. Moreover, the first automatised GPCR-tailored protocol able to connect the molecular with the subcellular level was developed. In KRc6.3 and 4 progress has been made towards a better understanding of the allosteric mechanisms of activations of CNS targets such as class A and B GPCRs, GABA and NGF, PI3Kα.		



Keywords:	SP8, SP2, SP6, CDP6, Allosteric regulation of proteins, Rational Drug Design, virtual high-throughput screening (vHTS), synaptic transmission, Ligand-gated Ion Channels, GPCRs, PI3K α , Molecular Dynamics simulations, cryptic binding sites.
Target Users/Readers:	Clinicians, computational neuroscience community, computational biophysics community, molecular biologists, computer scientists, biochemists, molecular modelling and simulations community, pharma companies, biotech companies, general public, HPC community, neuroimaging community, neuroinformaticians, researchers, scientific community, students.

Table of Contents

1. Overview	5
2. Introduction	5
3. KRc6.1 Allosteric modulation of pentameric ligand-gated ion channels	6
3.1 Outputs	6
3.1.1 Overview of Outputs	6
3.1.2 Output 1: Structural and dynamic characterisation of the GlyR physiological states by Molecular Dynamics	6
3.1.3 Output 2: PhD Thesis of Adrien CERDAN	7
3.1.4 Output 3: The GlyR Allosteric Ligand Library GRALL	7
3.1.5 Output 4: Set of allosteric ligands binding to nAChR	7
3.2 Validation and Impact	8
3.2.1 Actual and Potential Use of Output(s)	8
3.2.2 Publications	8
4. KRc6.2 Allosteric modulation of GPCRs	8
4.1 Outputs	8
4.1.1 Overview of Outputs	8
4.1.2 Output 1: Development of a molecular simulation tool able to characterise positive allosteric modulation of GPCRs	9
4.1.3 Output 2: Development of a Machine-learning-based pipeline for predicting ligand efficiency on GPCRs	9
4.1.4 Output 3: Application of Machine-learning-based approaches for predicting neuroprotective ligands	10
4.2 Validation and Impact	10
4.2.1 Actual and Potential Use of Output(s)	10
4.2.2 Publications	11
5. KRc6.3 Detecting druggable allosteric sites and hidden pockets	11
5.1.1 Overview of Outputs	11
5.1.2 Output 1: Understanding ligand binding selectivity of Class A GPCRs (Adenosine receptors)	12
5.1.3 Output 2: Understanding the activation mechanism of the A2a receptor (class A) and the glucagon receptor (class B) by means of enhanced sampling simulations.	12
5.1.4 Output 3: Understanding the influence of the E545K mutation in PI3K α dynamics and free energy landscape differences of active and inactive states induced by mutation E545K in PI3K α	13
5.1.5 Output 4: Structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF).	13
5.1.6 Output 5: Better understanding of cryptic binding sites (in CNS targets and in general) ...	14
5.2 Validation and Impact	14
5.2.1 Actual and Potential Use of Output(s)	14
5.2.2 Publications	14
6. KRc6.4 Targeting the protein-membrane interface for PI3Kα allosteric drug discovery	15
6.1 Outputs	15
6.1.1 Overview of Outputs	15

6.1.2	Output 1: Investigation of allosteric effects of the H1047R mutant contacting the cell membrane	16
6.1.3	Output 2: Development of ChemBioServer 2.0: a web-server for filtering, clustering and networking of chemical compounds, facilitating both drug discovery and repurposing	16
6.1.4	Output 3: Characterisation of the full-energy landscape of the WT and mutant PI3K α using metadynamics simulations	17
6.1.5	Output 4: Identification of novel modulators PI3K α E545K mutant	17
6.2	Validation and Impact	17
6.2.1	Actual and Potential Use of Output(s)	17
6.2.2	Publications	18
7.	Conclusion and Outlook	19

Table of Tables

Table 1: Output 1 and Output 2 Links	7
Table 2: Output 3 Links	7
Table 4: Output 1 Links	9
Table 5: Output 2 Links	10
Table 6: Output 1 Links	12
Table 7: Output 2 Links	12
Table 8: Output 3 Links	13
Table 9: Output 4 Links	13
Table 10: Output 5 Links	14
Table 11: Output 1 Links	16
Table 12: Output 2 Links	16
Table 13: Output 3 Links	17

Table of Figures

Figure 1: Allosteric regulation in therapeutic targets relevant for neuropathologies	1
--	---

History of Changes made to this Deliverable (post Submission)

Date	Change Requested / Change Made / Other Action
26 May 2020	Deliverable submitted to EC
31 Jul 2020	Deliverable approved by EC
2 Oct 2020	Minor editorial change by PCO
2 Oct 2020	Revised version resubmitted to EC by PCO via SyGMA

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 CDP focuses on three major classes of targets:

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

Unfortunately, due to the recent global reassessment of the Project, drug design activities will not be part of the next stage (SGA3).

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, and -omics data from other HBP WPs, are used to discover more effective modulators and help to design new drugs with enhanced selectivity and reduced side effects.

On the pentameric ligand-gated ion channels (pLGICs): we have characterised the physiological active state of the glycine receptor (GlyR) and completed μ s-long simulation trajectories for the resting, active, and desensitised state. Analysis of these trajectories has provided statistically relevant conformations of the GlyR modulatory sites in its physiological states, which is critical for structure-based drug discovery. The identification of 218 compounds from the literature with documented modulatory activity on GlyR function provided the first complete database (GRALL) for the allosteric modulation of one pLGIC. The GRALL library is freely distributed as a web-accessible database and is expected to be critically important for the development of more predictive theoretical and computational approaches for rational allosteric drug design. Finally, all the information above was used to initiate a virtual high-throughput campaign to prioritise compounds binding to a recently discovered modulatory site for the GlyR potentiation. *In vitro* testing of the most promising compounds by patch-clamp electrophysiology is ongoing.

On G-protein-coupled receptors (GPCRs): we have provided new tools to: 1) characterise allosteric regulation; 2) predict ligand efficiency by linking the molecular level data with the subcellular level; 3) find cryptic allosteric sites. We have also identified the main energetic bottlenecks for activation, clarified the activation mechanism of a class B receptor (the glucagon receptor) and made substantive progress towards understanding ligand selectivity in adenosine receptors. 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 α , a kinase dysregulated in many brain tumours, 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 α inhibition.

New methodologies were developed for drug-design, exploiting newly identified modulatory sites, as well as mechanisms of receptor's activation.

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

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

3.1 Outputs

3.1.1 Overview of Outputs

3.1.1.1 List of Outputs contributing to this KR

- Output 1: Structural and dynamic characterisation of the GlyR physiological states by Molecular Dynamics (C2885)
- Output 2: PhD Thesis of Adrien CERDAN (C2885)
- Output 3: The GlyR Allosteric Ligand Library GRALL (C2887)
- Output 4: Set of allosteric ligands binding to nAChR (C2888)

3.1.1.2 How Outputs relate to each other and the Key Result

The development of structure-based drug discovery approaches on brain receptors requires a structural description of the physiologically relevant states with atomic resolution, a collection of active compounds, and accurate and efficient strategies for library screening. In this respect, the knowledge obtained in Output 1, Output 2 and Output 3 will be instrumental to the development of structure-based approaches targeting the Glycine receptor. The combination of different simulation approaches including simplified binding free-energy calculations is critical for this Deliverable, SGA2 D8.6.2 (D53.2 D45). Output 4 should be considered as stand-alone.

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

Our characterisation of the GlyR active state by computational electrophysiology (Cerdan *et al.* Structure 2018, P1896) provided *bona fide* atomistic models for the resting, active, and desensitised states. Recently, our model of the active state has been challenged by another simulation study (Dämgen & Biggin, Structure 2020). Using computational electrophysiology, we have shown that the model of Dämgen & Biggin is non-physiological. Based on these observations and very recent cryo-EM reconstructions in native lipids we conclude that: (1) the open-channel structure we isolated in 2018 is physiologically relevant; (2) structural stability and water permeation in MD is not sufficient to assess the relevance of an open-channel structure; (3) computational electrophysiology is useful information for the structure/function annotation; (4) the GlyR conductance does not require the permeation of fully hydrated chloride ions, which suggests that partial dehydration might provide

the best compromise between ion conductance and selectivity. These results are part of manuscript that is currently under review in Structure.

3.1.3 Output 2: PhD Thesis of 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 (P2533). This thesis was partially funded by HBP.

Table 1: Output 1 and Output 2 Links

Component	Link to	URL
C2885	Model Repository	https://doi.org/10.17632/mh34bc6gty.1
	Technical Documentation	Atomistic model of the Glycine receptor (GlyR) active state isolated in simulation. https://data.mendeley.com/datasets/mh34bc6gty/1
	User Documentation	P1896 “An Ion-Permeable State of the Glycine Receptor Captured by Molecular Dynamics” Structure 11, 1555-1562 (2018). http://dx.doi.org/10.1016/j.str.2018.07.019

3.1.4 Output 3: The GlyR Allosteric Ligand Library GRALL

Starting from two recent reviews on the allosteric modulation of GlyR, we have collected 218 unique chemical entities with documented modulatory activity on GlyR and built a database named GRALL. (1) GRALL contains agonists, antagonists, positive and negative allosteric modulators, and several experimentally inactive compounds. (2) GRALL provides an unprecedentedly curated chemical database for allosteric modulation, which can be effectively exploited by chemo-informatics or machine learning approaches. (3) GRALL delivers an insightful view of the state of the art on the chemical modulation of GlyR; (4) the information contained in GRALL provides rigorous benchmarks for structure-based drug-discovery approaches, which is instrumental for future methodological development (component 2887). GRALL is freely distributed as a web-accessible database at: <https://ifm.chimie.unistra.fr/grall>. The companion paper has been published in Bioinformatics (P2442).

Table 2: Output 3 Links

Component	Link to	URL
C2887	Data Repository	https://ifm.chimie.unistra.fr/grall (in curation)
	Technical Documentation	P2442: Cerdan A, Sisquellas M, Pereira G, Barreto Gomes D, Changeux JP & Cecchini M. “The Glycine Receptor Allosteric Ligands Library (GRALL)”, Bioinformatics. http://dx.doi.org/10.1093/bioinformatics/btaa170
	User Documentation	See companion manuscript (P2442).

3.1.5 Output 4: Set of allosteric ligands binding to nAChR

The nicotinic acetylcholine receptor (nAChR) has been target of abundant drug discovery studies. During SGA2, we have found new targets for modulatory ligands with $\alpha 5$ and $\alpha 7$ subunits in mouse models, which provide original attempts towards the establishment of therapeutic strategies against schizophrenia and Alzheimer’s Disease. Also, we have explored the contribution of the subunit composition of oligomeric nAChRs to higher brain functions in the mouse and found that both $\alpha 7$ and $\beta 2$ subunits play an important role in the generation of ultraslow fluctuations in the prefrontal cortex and that the $\beta 2$ subunit is required for synchronised activity patterns. Also, we found that chronic application of the nicotinic drug mecamylamine disrupts the generation of ultraslow fluctuations *in vivo*. These experiments underline the discovery of new effect of nicotinic ligands acting upon AChRs oligomers, which modulate activity patterns characteristic of conscious access (P1467).

3.2 Validation and Impact

3.2.1 *Actual and Potential Use of Output(s)*

The coordinates of the GlyR active state from Molecular Dynamics have been published and deposited in PLUS. The trajectories of the resting and desensitised states will be made available. These models provide the first complete set of structures for the main physiological states of the same neurotransmitter receptor. The availability of the first complete dataset of small-molecule modulators of GlyR will have a strong impact on the scientific community providing quantitative benchmark to test computational strategies for allosteric drug design. The structure-based screening campaign we have initiated led to the prioritisation of several promising candidates for GlyR potentiation, which wait for experimental testing. All the above is expected to contribute to the establishment of state-based pharmacology approaches, which holds the potential to revolutionise the way allosteric drug design is conceived.

3.2.2 *Publications*

Output 2: (P2533) Cerdan A “Exploring synaptic transmission and regulation in ionotropic receptors by molecular dynamics and computational electrophysiology”, (2019).
<http://www.theses.fr/2019STRAF006/abes>

Output 3: (P2442) Cerdan A, Sisquellas M, Pereira G, Gomes DEB, Changeux JP & Cecchini M. “The Glycine Receptor Allosteric Ligands Library (GRALL)”, Bioinformatics (2020).
<https://doi.org/10.1093/bioinformatics/btaa170>

This paper reports on the first complete database of positive and negative chemical modulators of the Glycine receptor channel.

Output 4: (P2351 - in validation process) Oasa S, Changeux JP et al “A strategy for designing allosteric modulators of transcription factor dimerization”, PNAS 117, 2683-2686 (2020).
<https://doi.org/10.1073/pnas.1915531117>

This paper describes a promising approach toward drug development targeting transcription factors over activity and protein-protein interaction more generally.

4. KRc6.2 Allosteric modulation of GPCRs

Contributors: Giulia ROSSETTI (SP2 and SP8 - JUELICH, P20) and Paolo CARLONI (SP6 and SP8 - JUELICH, P20)

4.1 Outputs

4.1.1 *Overview of Outputs*

4.1.1.1 List of Outputs contributing to this KR

- Output 1: Development of a molecular simulation tool able to characterise positive allosteric modulation of GPCRs (C2886; C2893 - in T8.6.2)
- Output 2: Development of a Machine-learning-based pipeline for predicting ligand efficiency on GPCRs (C2894 - in T8.6.2 and T2.3.8)

4.1.1.2 How Outputs relate to each other and the Key Result

Output 1 is a direct consequence of what we developed in M1-M12 (C2890) and it is fundamental to detect and quantify allosteric effects in GPCRs. This knowledge is key to exploit allosterism as a tool for GPCRs modulation. The knowledge obtained in Output 1 can be used as an input for Output 2. The latter is the natural follow-up of what we achieved in M1-M12 (C2889).

The combination of Output 1 and Output 2 is fundamental for the current CDP6 SGA2 M24 Deliverable D8.6.2 (D53.2 D45), i.e. to develop of new HPC approaches for the rational optimisation of drugs.

4.1.2 *Output 1: Development of a molecular simulation tool able to characterise positive allosteric modulation of GPCRs*

Allosteric modulation is involved in a plethora of diverse protein functions, which are fundamental for cells' life. This phenomenon can be thought as communication between two

topographically distinct site of a protein structure. How this communication occurs is still matter of debate. Many different descriptions have been presented so far. Here we consider a specific case where any conformational change is involved upon allosteric modulator binding and the phenomenon is depicted as a vibrational energy diffusion process between distant protein regions. We applied this model, by employing computational tools, to the human muscarinic receptor M2, a transmembrane protein G-protein coupled receptor known to undergo allosteric modulation whose recently X-ray structure has been recently resolved both with and without the presence of a particular allosteric modulator. Our MD based tool, performed on these two receptor structures, showed that the allosteric modulator modifies the vibrational energy current between functionally relevant regions of the protein; this allows to identify the main residues responsible for this modulation. These results contribute to shed light on the molecular basis of allosteric modulation and may help design new allosteric ligands. See in SGA2 Deliverable D8.6.1 (M1-M12), the publication P1544 (Maggi *et al.* J. Phys. Chem. Lett. 2018) and the released component C2890. In the current D6.6.2 (M12-M24), component release: C2886, C2893 and P2321 (Maggi *et al.* Scientific Reports, 2020).

Table 3: Output 1 Links

Component	Link to	URL
C2886	Data Repository	https://rwth-aachen.sciebo.de/s/QtOrUchGw7v9Z08 (in curation)
	Technical Documentation	https://rwth-aachen.sciebo.de/s/QtOrUchGw7v9Z08
	User Documentation	doi:10.1038/s41598-020-59289-5
C2893	Model Repository	https://collab.humanbrainproject.eu/#/collab/45324/nav/311401
	Technical Documentation	http://plumed.github.io/doc-master/user-doc/html/
	User Documentation	http://plumed.github.io/doc-master/user-doc/html/

4.1.3 *Output 2: Development of a Machine-learning-based pipeline for predicting ligand efficiency on GPCRs*

Pharmacokinetics/dynamics studies are among the most complex tasks to accomplish during drug discovery and development. Under the umbrella of systems biology and systems pharmacology, pharmacodynamic models have evolved from empirical descriptions to mathematical models that describe complex biological systems. However, a deep understanding of the molecular features underlying drug-protein interactions is fundamental to improve the mathematical model definition. Thus, we are developing an automatic structural systems biology approach to bridge these two levels of resolution.

As an application, we have studied the effect drug-adenosine A2A receptor interactions on the metabolic pathway. For this, we adapted the metabolic mathematical model proposed by Kotalesky *et al.* in SP6. We were able to reproduce the dose-response curves for three known agonists, i.e. Adenosine, Regadenoson and NECA. Moreover, our protocol allows the characterisation of the receptor variants effects on the pathway. In particular, we have analysed the effects of several site-directed mutagenesis experiments using NECA as agonist. The results were validated against experimental data. Finally, we have also included a Machine-Learning based discriminator, able to analyse and ensemble of unknown ligands, distinguish between agonist/antagonist-like features and select the respective metabolic model. This will allow to predict the effect of possible drug molecules or molecular modulators on GPCRs. See in D8.6.1 (M1-M12) publication P1557 (Cao *et al.* Molecules, 2018) and the released component C2889, and in the current D8.6.2 (M13-M24), P2082 (Rossetti *et al.* Biochemical Society Transactions, 2019) and C2894.

Table 4: Output 2 Links

Component	Link to	URL
C2894	Data Repository	P2082: https://juser.fz-juelich.de/record/866070 (in curation)
	Technical Documentation	P2082: https://doi.org/10.1042/BST20190048
	User Documentation	P2082: Rossetti, G, Kless, A, Lai, L, Outeiro, TF, P Carloni, "Investigating targets for neuropharmacological intervention by molecular dynamics simulations, Biochemical Society Transactions, 2019" Biochemical Society Transactions, Vol. 47, No. 3, (2019).

4.1.4 ***Output 3: Application of Machine-learning-based approaches for predicting neuroprotective ligands***

Polyglutamine (polyQ) diseases, including Huntington's disease, are characterised by an expansion of cytosine-adenine-guanine (CAG) repeats in the coding region of the disease-linked genes. As CAG codes for glutamine, the disease-linked proteins contain a polyQ tract, eponymous for these diseases.

Onset of these diseases is typically in midlife and coincides with progressive neuronal dysfunction and neuron loss. To date, polyQ diseases remain incurable. Approved treatments are limited to symptomatic relief, whereas the progression of neuronal decline is unaffected. Therefore, there is a clear need for neuroprotective strategies to delay disease onset and/or progression.

TRMT2a was previously identified as a strong modifier of polyQ-induced toxicity in an unbiased large-scale screen in flies. TRMT2a is well conserved between fly and human. Hence, we attempted to identify TRMT2a inhibitors by using computational methods including machine learning and molecular simulations and, we identified a putative cryptic site, potentially able to allosterically affect RNA binding. Inhibitors were tested in cell-based assays. These showed that several compounds capable to decrease polyQ-induced toxicity and reduce polyQ aggregation. The results were exploited for the (Confidential) Patent application PT 1.2833 EP, ROSSETTI Giulia, Michael A. MARGREITER, Jörg SCHULZ, Aaron VOIGT, Dierk NIESSING, Monika WITZENBERGER, John SHAH. "TRMT2A INHIBITORS FOR USE IN THE TREATMENT OF POLYGLUTAMINE DISEASES".

4.2 Validation and Impact

4.2.1 ***Actual and Potential Use of Output(s)***

Output 1 and 2 are currently being used to develop allosteric modulators against GPCRs in SP8. Moreover, they are also connecting algorithms and pipelines developed in SP6 to applications in SP2.

For instance, the metabolic mathematical model proposed by Kotalesky *et al.* in SP6 are used in CDP6 to understand and predict GPCRs' allosteric modulation in SP8 and SP2.

Moreover, the tools and protocols developed here will be implemented in EBRAINS (during SGA3) as a molecular platform. Namely, this will allow the prediction of the effects of neuroactive ligands at the subcellular level both for wild-type proteins as well as aberrant neurological cascades associated with disease-linked mutations.

The planned, innovative activity is of interest to industry partners active in the fields of neurobiology and neuropharmacology: Indeed, Merck, Sanofi, Ernesto Illy Foundation and NBD (Nostrum Biodiscovery) have offered their support for SGA3, and two of them have substantiated their support with co-funding.

4.2.2 Publications

Output 1: (P2321) Maggi, L, P Carloni, and G Rossetti. "Modeling the Allosteric Modulation on a G-Protein Coupled Receptor: the Case of M2 Muscarinic Acetylcholine Receptor in Complex with LY211960." *Scientific Reports* 10, no. 1: 1-9 (2020).

doi:10.1038/s41598-020-59289-5.

This paper shows that the allosteric modulator modifies the vibrational energy current between functionally relevant regions of the protein; this allows to identify the main residues responsible for this modulation. These results contribute to shed light on the molecular basis of allosteric modulation and may help design new allosteric ligands.

Output 2: (P2082) Rossetti, Giulia, Achim Kless, Luhua Lai, Tiago F Outeiro, and Paolo Carloni. "Investigating Targets for Neuropharmacological Intervention by Molecular Dynamics Simulations." *Biochemical Society Transactions* 40 BST20190048-10 (2019).

<https://doi.org/10.1042/BST20190048>

This paper is a short review of all the Molecular-Simulations-based approaches, developed in HBP, that had an impact in developing new strategies for neuropharmacology. Notably, a Pharma Company member, Achim KLESS, was among the co-authors.

Output 3: (Confidential) Patent application PT 1.2833 EP, Rossetti Giulia, Michael A. Margreiter, Jörg Schulz, Aaron Voigt, Dierk Niessing, Monika Witzemberger, John Shah. "TRMT2A INHIBITORS FOR USE IN THE TREATMENT OF POLYGLUTAMINE DISEASES"

The present invention relates to TRMT2A inhibitors for use in the treatment and/or the prevention of polyglutamine diseases.

5. KRc6.3 Detecting druggable allosteric sites and hidden pockets

Contributors: Francesco Luigi GERVASIO (SP8 - UCL, P82).

5.1.1 Overview of Outputs

5.1.1.1 List of Outputs contributing to this KR

- Output 1: Understanding ligand binding selectivity of Class A GPCRs (Adenosine receptors)
- Output 2: Understanding the activation mechanism of the A2a receptor (class A) and the glucagon receptor (class B) by means of enhanced sampling simulations.
- Output 3: Understanding the influence of the E545K mutation in PI3K α dynamics and free energy landscape differences of active and inactive states induced by mutation E545K in PI3K α .
- Output 4: Structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF).

- Output 5: Better understanding of cryptic binding sites (in CNS targets and in general).

5.1.1.2 How Outputs relate to each other and the Key Result

Output 1 is fundamental to understand selectivity in GPCRs. Output 2 is about understanding the mechanism of activation of class A and class B GPCRs, of fundamental importance for a rational allosteric regulation of their activity. The knowledge from Outputs 1, 2, 4 and 5 are key for the rational design of selective biased ligands and thus to obtain safer drugs with less undesirable effects. The same computational techniques developed in Output 1-2 are then used in Output 3 to understand the influence of the E545K mutation in PI3K α dynamics and the structural and dynamical characterisation of the Nerve growth factor precursor (pro-NGF) in Output 4.

5.1.2 *Output 1: Understanding ligand binding selectivity of Class A GPCRs (Adenosine receptors)*

To understand the molecular determinants behind ligand binding selectivity in a prototypical class A family receptor, the adenosine receptor, by means of enhanced sampling techniques. The extensive structural similarity of adenosine receptors complicates the design of selective ligands. The problem of selective targeting is a general concern in GPCRs, and in this respect adenosine receptors are a prototypical example. Using enhanced sampling simulations, we decipher the determinants of selectivity of ligands in A2a and A1 adenosine receptors. Our model shows how small differences in the binding pocket and in the water network around the ligand can be leveraged to achieve selectivity.

Table 5: Output 1 Links

Component	Link to	URL
C3002 release	Data Repository	https://www.plumed-nest.org/eggs/19/048/
	Technical Documentation	https://www.plumed-nest.org/eggs/19/048/
	User Documentation	https://www.plumed-nest.org/eggs/19/048/

5.1.3 *Output 2: Understanding the activation mechanism of the A2a receptor (class A) and the glucagon receptor (class B) by means of enhanced sampling simulations.*

To calculate the relative populations and energy barriers separating the conformational repertoire of the adenosine A_{2a} and glucagon receptors, respectively a class A and class B GPCRs. We derived an activation collective variable from molecular dynamics simulations of the receptors, and reconstructed the free energy landscape for the receptor activation from a 12/16 replicas well-tempered-ensemble, parallel-tempering Metadynamics simulations. The study finding provides valuable information for rational drug development (Mattedi *et al.* 2020 under review).

Table 6: Output 2 Links

Component	Link to	URL
C3002 release	Data Repository	https://www.plumed-nest.org/eggs/20/006/
	Technical Documentation	https://www.plumed-nest.org/eggs/20/006/
	User Documentation	https://www.plumed-nest.org/eggs/20/006/

5.1.4 ***Output 3: Understanding the influence of the E545K mutation in PI3Ka dynamics and free energy landscape differences of active and inactive states induced by mutation E545K in PI3Ka.***

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. 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 PI3K α , 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.

Table 7: Output 3 Links

Component	Link to	URL
C3003 release "PI3Ka-dataset"	Data Repository	Uploaded in E-brains with the help of the curation team https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802
	Technical Documentation	https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802
	User Documentation	https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802

5.1.5 ***Output 4: 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 allowed us to gain novel insights on the prodomain, both in isolation and in the context of proNGF.

Table 8: Output 4 Links

Component	Link to	URL
C3002 release	Data Repository	https://doi.org/10.6084/m9.figshare.6866453
	Technical Documentation	https://doi.org/10.1016/j.str.2018.09.013
	User Documentation	https://doi.org/10.1016/j.str.2018.09.013

5.1.6 *Output 5: Better understanding of cryptic binding sites (in CNS targets and in general).*

Cryptic (hidden) allosteric binding sites might provide outstanding opportunities to target receptors that are otherwise “undruggable”. We studied the mechanism of opening of a number of cryptic sites, including in CNS targets such as GPCRs and GABA. This allowed us to further develop effective computational tools such as SWISH (C3002) to predict and model cryptic site opening.

Table 9: Output 5 Links

Component	Link to	URL
C3002	Data Repository	https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/blob/master/kras_swish.tar.gz
	Technical Documentation	https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/blob/master/kras_swish.tar.gz
	User Documentation	https://github.com/Gervasiolab/Gervasio-Protein-Dynamics/blob/master/kras_swish.tar.gz

5.2 Validation and Impact

5.2.1 *Actual and Potential 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 (Output 1-2). The reconstructed free energy landscapes for the activation of the class B receptors are useful to understand their conformational repertoire as the on/off paradigm seems to be not representative of the real biophysics of the system (Output 2). Moreover, the extensive structural similarity among GPCR receptors complicates the design of selective ligands, increasing undesirable effects in the patient. The problem of selective targeting is a general concern in GPCRs, our results contribute towards the understanding of the molecular determinants for binding affinity and selectivity, and will hopefully aid the design of more selective and safer drugs (Output 1).

Studying the free energy of the PI3K α 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 (Output 3).

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 (Output 4).

A better understanding of the nature of cryptic binding sites in general and in CNS targets might open many opportunities for the rational design of modulators for difficult or even “undruggable” targets (output 5).

5.2.2 *Publications*

Output 1: (P2089) Mattedi G, Deflorian F, Mason J S, de Graaf C, Gervasio F L. Understanding ligand binding selectivity in a prototypical GPCR family. *Journal of chemical information and modeling*, 59, 2830-2836, 2019. <https://doi.org/10.1021/acs.jcim.9b00298>

This paper shows the molecular determinants behind ligand selectivity in the class A prototypical family of adenosine receptors. These results contribute to shed light on the molecular basis of ligand selectivity and may help design safer drugs targeting GPCRs.

Output 3: (P2401) Galdadas I, Gervasio F L and Cournia Z. Unravelling the effect of the E545K mutation on PI3K α kinase. Chemical Science 2020 Accepted. <https://doi.org/10.1039/C9SC05903B>

In this work, we used biased and unbiased Molecular Dynamics simulations to uncover, for the first time, the free energy landscape of the wild type and E545K PI3K α , which is a hotspot mutation found in many cancer types. We reveal the mechanism by which the oncogenic mutation E545K leads to PI3K α activation in atomic-level detail, and show that its mechanism of action is considerably more complex than previously thought.

6. KRc6.4 Targeting the protein-membrane interface for PI3K α allosteric drug discovery

Contributor: Zoe COURNIA (SP8 - BRFAA, P120)

6.1 Outputs

6.1.1 Overview of Outputs

In the context of KRc6.4 a rational, structure-based methodology has been devised, which identifies small molecules that can modulate protein-membrane interactions. This methodology was applied to identify novel PI3K α allosteric inhibitors against mutated PI3K α , implicated in glioblastoma. (C3003)

For C3003 we describe a) a new methodology developed for the discovery of protein-membrane interfaces using Machine Learning, b) a new methodology for filtering, clustering and networking of chemical compounds, c) the full-energy landscape of the WT and mutant E545K PI3K α and d) novel inhibitors of PI3K α E545K mutant using computational screening, *in vitro* and *in vivo* assays.

6.1.1.1 List of Outputs contributing to this KR

- Output 1: Methodology development for the discovery of protein-membrane interfaces using Machine Learning
- Output 2: Development of ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds
- Output 3: Characterisation of the full-energy landscape of the WT and mutant E545K PI3K α using metadynamics simulations
- Output 4: Identification of novel modulators PI3K α E545K mutant.

6.1.1.2 How Outputs relate to each other and the Key Result

Outputs 1 and 2 are connected to KRc6.4 to create a workflow targeting the protein-membrane interface with small molecules. The methodology of Output 1 & 2 was applied to Output 3, which was used to deliver Output 4, i.e. the successful development of novel small molecule inhibitors of the mutant E545K PI3K α involved in glioblastoma. All four outputs are connected to C3003 and each output has been described in a corresponding C3003 release, which is described below.

6.1.2 *Output 1: Investigation of allosteric effects of the H1047R mutant contacting the cell membrane*

A PI3K α hotspot mutant, H1047R, which is prevalent in brain, breast, and other cancers, has been shown to act by modulating the way that the kinase interacts with the cell membrane. We performed dynamical network analysis of WT and H1047R mutant trajectories reported in <https://repo.vi-seem.eu/handle/21.15102/VISEEM-173> and found that H1047 is connected to the PI3K α membrane-binding loop (residues 863-873) through a pathway of ten residues, while the mutant 1047R significantly shortens this path and connects itself to the membrane-binding loop through a pathway of six residues. (C3003)

Table 10: Output 1 Links

Component	Link to	URL
C3003 release "H1047R- membrane"	Data Repository	https://repo.vi-seem.eu/handle/21.15102/VISEEM-173
	Technical Documentation	http://dx.doi.org/10.5281/zenodo.3701442
	User Documentation	http://dx.doi.org/10.5281/zenodo.3701442

Output 1 has been released in C3003 as "H1047R-membrane".

6.1.3 *Output 2: Development of ChemBioServer 2.0: a web-server for filtering, clustering and networking of chemical compounds, facilitating both drug discovery and repurposing*

ChemBioServer provides researchers the ability to (i) browse and visualise compounds along with their physicochemical and toxicity properties, (ii) perform property-based filtering of chemical compounds, (iii) explore compound libraries for lead optimisation based on perfect match substructure search, (iv) re-rank virtual screening results to achieve selectivity for a protein of interest against different protein members of the same family, (v) perform clustering among the compounds based on their physicochemical properties providing representative compounds for each cluster, (vi) construct and visualise a structural similarity network of compounds providing a set of network analysis metrics, (vii) combine a given set of compounds with a reference set of compounds into a single structural similarity network providing the opportunity to infer drug repurposing due to transitivity, (viii) remove compounds based on their similarity with unwanted/toxic moieties and (ix) build custom compound mining pipelines. (C3003)

Table 11: Output 2 Links

Component	Link to	URL
C3003 Release "ChemBioSer ver".	Service	https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,ChemBioServer
	Technical Documentation	https://collab.humanbrainproject.eu/#/collab/19/nav/2108?state=software,ChemBioServer
	User Documentation	doi: 10.1093/bioinformatics/btz976

Output 2 has been released in C3003 as "ChemBioServer".

6.1.4 **Output 3: Characterisation of the full-energy landscape of the WT and mutant PI3K α using metadynamics simulations**

We performed biased and unbiased Molecular Dynamics simulations of PI3K α and revealed the mechanism by which E545K leads to PI3K α activation in atomic-level detail, which is considerably more complex than previously thought. Our simulations show two different displacement paths resulting from the E545K mutation: the detachment of the nSH2 domain from the helical domain, and the sliding along the protein surface. The perturbation around E545K is propagated to neighbouring domains leading to “unlocking” the activation-loop and allowing it to adopt active-like conformations. (C3003)

Table 12: Output 3 Links

Component	Link to	URL
C3003 release “PI3K α - dataset”	Data Repository	Uploaded in E-brains with the help of the curation team https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802
	Technical Documentation	https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802
	User Documentation	https://kg.ebrains.eu/search/instances/Model/7f44abeb3068cdc74506bac6e72a8802

Output 3 has been released in C3003 as “PI3K α -dataset”.

6.1.5 **Output 4: Identification of novel modulators PI3K α E545K mutant**

Using data from the simulations, representative conformations for the PI3K α E545K were exploited to identify allosteric pockets for selective inhibition of the mutant protein. Identified binding sites were examined for allosteric communication with the active (ATP) site of PI3K α and were further used in computer-aided drug design. Subsequent post-processing of top-scoring compounds for toxicity, pose viability, chemical diversity, and physicochemical properties led to identifying initial lead compounds. PI3K α activity assays on these compounds were performed using PIP2-containing liposomes. Assays showed that two compounds, BRF37 and BRF41 were potent selective inhibitors of PI3K α E545K mutant. In order to test the *in vivo* activity of the compounds we used immunodeficient mice with heterotopic xenografts after subcutaneous injection of cells from the colorectal carcinoma cell line DLD1 (WT PI3K α) and an isogenic colorectal carcinoma cell line DLD1 (E545K mutant PI3K α). Compounds BRF37 and BRF41 were shown to be highly specific in mice, in the inhibition of the action of the mutant PI3K α . (C3003)

The novel inhibitors were verified *in vitro* and *in vivo* and a patent application was deposited: Greek OBI Patent Application number 22-0002278706 and German Patent Application No. 10 2020 202 356.5 (A PCT application will be deposited in 12 months).

Output 4 has been released in C3003 as “Patent-application2-E545K”.

6.2 Validation and Impact

6.2.1 **Actual and Potential Use of Output(s)**

Outputs 1 and 2 are part of our methodology to create virtual screening workflows targeting the protein-membrane interface with selective inhibitors. Output 3 led to the mechanism by which E545K leads to PI3K α activation in atomic-level detail. Output 4 successfully developed novel small molecule inhibitors of the mutant E545K form of PI3K α . These inhibitors are the first allosteric

regulators reported for PI3K α E545K mutant. Results obtained in Output 1, 2, 3 and 4 have the potential to translate into targeted therapeutics.

6.2.2 Publications

Output 1: (P2538 - in press - SGA3 publication) Cournia Z, Chatzigoulas A. Allostery in membrane proteins. Doi: <http://dx.doi.org/10.5281/zenodo.3701442>.

This paper summarises recent advances in membrane protein allostery and analyses allosteric contacts of H1047R with the membrane.

Output 2: (P2352) Karatzas E, Zamora JE, Athanasiadis E, Dellis D, Cournia Z, Spyrou GM. ChemBioServer 2.0: An advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing. *Bioinformatics*. 2020 pii: btz976. doi: 10.1093/bioinformatics/btz976.

This paper describes “ChemBioServer 2.0”, a free web-server for filtering, clustering and networking of chemical compound libraries facilitating both drug discovery and repurposing.

Output 3: (P2401) I Galdadas, FL Gervasio, Z Cournia. Unravelling the effect of the E545K mutation on PI3K α kinase. *Chemical Science*, 2020, doi:10.1039/C9SC05903B

This paper describes MD simulations of PI3K α and uncovers, for the first time, the mechanism by which E545K leads to PI3K α activation in atomic-level detail, which is considerably more complex than previously thought.

Output 4: (Confidential) Z. Cournia, S Christoforidis, A Papafotika, V-E Karamani, A Tzacos, E. Vrettos, C. Chatzigiannis, A Efstratiadis, A Klinakis. “Method of preparation and use of phosphoinositide 3-kinase inhibitors in treating cancer”. OBI Patent Application number 22-0002278706 and German Patent Application No. 10 2020 202 356.5.

The invention discloses the first allosteric inhibitors of PI3K α selective for the mutant E545K PI3K α , 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.

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 to explore receptor's activation with atomic resolution are being developed. The information collected, along with the developed methodologies, were exploited in M24. Namely, for KRc6.1, on the Glycine receptor, which is an important drug target for chronic inflammatory pain, we have initiated the first virtual high-throughput screening campaign targeting a modulatory site for potentiation. This campaign relies on: (1) the structural characterisation of the physiologically relevant states of the receptor, which was realised by an original combination of Molecular Dynamics simulations and computational electrophysiology; (2) the identification of statistically relevant conformations of the modulatory sites in the various states of the receptor, which was obtained from the analysis of μ s-long MD trajectories; and (3) the collection of a large number of experimentally active and inactive compounds on GlyR including positive and negative modulators, which provided an original benchmark for virtual screening strategies; (4) the development of accurate and efficient strategies for *in silico* screening based on conformer enumeration of ligands prior to rigid docking. By screening the French National Chemical Library using this protocol, several new scaffolds for GlyR potentiation have been prioritised, which will be tested *in vitro* during SGA3. For KRc6.2, capitalising on the molecular information collected not only in SP8, but also in SP2 and SP6, we have initiated the first virtual high-throughput screening against selected classes of GPCRs, relevant for neuropathologies, implementing kinetics and thermodynamics information. Moreover, the first automatised GPCR-tailored protocol able to connect the molecular with the subcellular level was developed. This allowed for the first time, to predict the effect of ligand binding at subcellular level. Also a patent application (PT 1.2833 EP) has been released. For KRc6.3 and KR6.4 substantive progress has been made towards a better understanding of the allosteric mechanisms of activations of CNS targets such as class A and B GPCRs, GABA and NGF, PI3K α . Moreover, we have gained deeper knowledge on the mechanism of formation of cryptic allosteric sites both in CNS targets and in general. This scientific, conceptual, and technological progress has already been applied for KRc6.4, for the study and discovery of novel PI3K α inhibitors specific for the H1047R or E545K mutations. The novel inhibitors were verified *in vitro* and *in vivo* and two patent applications were deposited (first patent application: PCT/EP2019/072648, second patent application: OBI Patent Application number 22-0002278706 and German Patent Application No. 10 2020 202 356.5 (A PCT application will be deposited in 12 months).