White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Leeds.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/143004/

Version: Accepted Version

Article:
Wodak, SJ, Paci, E, Dokholyan, NV et al. (2019) Allostery in Its Many Disguises: From Theory to Applications. Structure, 27 (4). pp. 566-578. ISSN: 0969-2126

https://doi.org/10.1016/j.str.2019.01.003

© 2019 Published by Elsevier Ltd. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Allostery in its many disguises: from theory to applications

Shoshana J. Wodak[1], Emanuele Paci[2], Nikolay V. Dokholyan[3a,3b], Igor N. Berezovsky[4], Amnon Horovitz[5], Jing Li[6], Vincent J. Hilser[6], Ivet Bahar[7], John Karanicolas[8], Gerhard Stock[9], Peter Hamm[10], Roland H. Stote[11], Jerome Eberhardt[11], Yassmine Chebaro[11], Annick Dejaegere[11], Marco Cecchini[12], Jean-Pierre Changeux[13], Peter G. Bolhuis[14], Jocelyne Vreede[14], Pietro Faccioli[15], Simone Orioli[15], Riccardo Ravasio[16] Le Yan[17], Carolina Brito[18], Matthieu Wyart[16], Paraskevi Gkeka[19], Ivan Rivalta[20], Giulia Palermo[21], J. Andrew McCammon[21], Joanna Panecka-Hofman[22], Rebecca C. Wade[23], Antonella Di Pizio[24], Masha Y. Niv[25], Ruth Nussinov[26a,26b], Chung-Jung Tsai[26a], Hyunbum Jang[26a], Dzmitry Padhorny[27], Dima Kozakov[27], Tom McLeish[28]*

[1]*VIB-VUB Centre for Structural Biology, Brussels, Belgium*
[2]*Astbury Centre, University of Leeds, UK*
[3a]*Department of Biochemistry & Biophysics, University of North Carolina at Chapel Hill, NC, USA*
[3b]*Departments of Pharmacology and Biochemistry & Molecular Biology, Penn State Medical Center, Hershey, PA, USA*
[4]*Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), and Department of Biological Sciences, National University of Singapore*
[5]*Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel*
[6]*Departments of Biology and T.C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, USA*
[7]*School of Medicine, University of Pittsburgh, USA*
[8]*Fox Chase Cancer Center*, *Philadelphia, USA*
[9]*Biomolecular Dynamics, Institute of Physics, Albert Ludwigs University, Freiburg, Germany*
[10]*Department of Chemistry, University of Zurich, Switzerland*
[11]*Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), Illkirch, France*
[12]*Institut de Chimie de Strasbourg, UMR7177 CNRS & Université de Strasbourg, France*
[13]*Institut Pasteur & Collège de France, Paris, France*
[14]*van 't Hoff Institute for Molecular Sciences (HIMS), University of Amsterdam, Netherlands*
[15]*Physics Department, Università di Trento and INFN-TIFPA*
[16]*Institute of Physics, Ecole Polytechnique Fédérale de Lausanne, Switzerland*
[17]*Kavli Institute for Theoretical Physics, University of California, Santa Barbara, CA 93106*
[18]*Instituto de Física, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil*
[19]*Structure Design and Informatics, Sanofi R&D, Chilly-Mazarin, France*
[20]*École Normale Supérieure de Lyon, Université de Lyon, CNRS, Université Claude Bernard Lyon 1, France*
[21]*Department of Chemistry and Biochemistry, University of California, San Diego, USA*
[22]Division of Biophysics, Institute of Experimental Physics, Faculty of Physics, University of Warsaw, Poland
[23]*Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS) and Center for Molecular Biology (ZMBH), DKFZ-ZMBH Alliance, Heidelberg University, Germany; Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany*
[24]*Leibniz-Institute for Food Systems Biology, Technical University of Munich, Germany*
[25]*Institute of Biochemistry, Food Science and Nutrition, Robert H Smith Faculty of Agriculture Food and Environment, The Hebrew University, Israel*
[26a]*Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, USA,*

[26b]*Sackler Institute of Molecular Medicine Department of Human Genetics and Molecular Medicine Sackler School of Medicine, Tel Aviv University, Israel*
[27]*Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794*
[28]*Department of Physics, University of York, United Kingdom*

*Corresponding Author:
Shoshana J. Wodak
Email: Shoshana.Wodak@gmail.com

Abstract

Allosteric regulation plays an important role in many biological processes, such as signal transduction, transcriptional regulation and metabolism. Allostery is rooted in the fundamental physical properties of macromolecular systems, but its underlying mechanisms are still poorly understood. A collection of contributions to a recent interdisciplinary CECAM (Centre Européen de Calcul Atomique et Moléculaire) workshop is used here to provide an overview of the progress and remaining limitations in the understanding of the mechanistic foundations of allostery gained from computational and experimental analyses of real proteins systems and model systems. The main conceptual frameworks instrumental in driving the field are discussed. We illustrate the role of these frameworks in illuminating molecular mechanisms and explaining cellular processes, and describe some of their promising practical applications in engineering molecular sensors and informing drug design efforts.

Introduction

Allostery refers to processes whereby a binding event at one site of a biological macromolecule affects the binding activity at another distinct functional site, enabling the regulation of the corresponding function. Since its initial formulations over 50 years ago (Changeux, 1961, 2011; Koshland et al., 1966; Monod and Jacob, 1961; Monod et al., 1965), allosteric regulation has been recognized as playing a key role in many biological processes, most prominently in signal transduction (Changeux, 2012; Changeux and Edelstein, 2005; Falke and Piasta, 2014; Nussinov et al., 2013), molecular machine function (Saibil, 2013), transcriptional regulation (Li et al., 2017; Wright and Dyson, 2015), and metabolism (Link et al., 2014). Allostery is rooted in the fundamental physical properties of macromolecular systems, and probably of other materials as well. However, the detailed mechanisms whereby these physical properties underpin allostery are not fully understood. Furthermore, allosteric effects are modulated by the cellular context in both health and disease.

Computational approaches have all along played an important role in the investigation of allosteric mechanisms. They have provided insights into some of the underpinnings of allostery (Dokholyan, 2016; Guo and Zhou, 2016; Schueler-Furman and Wodak, 2016) and have recently shown great promise in various practical applications, such as engineering regulatory modules in proteins and identifying allosteric binding sites that can be targeted by specific drugs. Notable examples of the latter application include re-sensitizing resistant hepatitis C variants by a combination therapy that involves binding to the allosteric site of NS5A (Sun et al., 2015), allosteric inhibitors of HIV integrase (Hayouka et al., 2007), or the

discovery of allosteric drugs that inhibit PARP-1 without hampering its action in cancer-related DNA repair deficiencies (Steffen et al., 2014).

One should also mention various recent bioinformatics approaches, which analyse sequence information (patterns of sequence conservation or correlated mutations) with the goal of uncovering signals of evolutionary pressure that may either inform or validate mechanistic aspects of allosteric processes (Dima and Thirumalai, 2006; Kass and Horovitz, 2002; Livesay et al., 2012; Lockless and Ranganathan, 1999; May et al., 2007). Here, too, the vast increase in available data on protein sequences from different organisms and massive data on human polymorphism derived from next generation sequencing efforts (Clarke et al., 2016) is providing unprecedented (and still largely untapped) opportunities for investigating the role of evolution in shaping allosteric regulation.

A recent CECAM (Centre Européen de Calcul Atomique et Moléculaire) workshop brought together about 30 computational biophysicists, protein modellers and bioinformaticians as well as experimentalists for an inspiring 2.5 days of stimulating talks and discussions. Among the important topics addressed were the new insights gained into the mechanistic foundations of allostery from computational and experimental analyses of real protein systems, as well as from very simple *in silico* toy materials. Also presented were informative examples describing how allostery enables information processing in cellular signalling cascades. Real excitement was generated by reports on the rational design of allosteric systems that can be modulated to produce desired activity and cellular behaviour, or engineered to act as sensitive molecular sensors. Encouraging results were also

described on the rational discovery of allosteric drugs by combining computational and experimental approaches.

In the following we summarize the highlights of the meeting. Further details are provided in the Supplementary Material.

Mechanistic underpinnings of allostery: insights from computational and experimental approaches.

The current understanding of allosteric systems has been increasingly influenced by the so-called ensemble model of allostery (Hilser et al., 2012; Motlagh et al., 2014), itself rooted in the seminal model of Monod Wyman and Changeux (MWC) (Monod et al., 1965), derived from studies on hemoglobin (Perutz, 1970), the 'ancestor' of all allosteric systems. According to the ensemble model, first described in the eighties (Cooper, 1984; Frauenfelder et al., 1988), the allosteric behavior of a macromolecular system arises from the properties of the native free energy landscape of the system, and how this landscape is remodeled by various 'perturbations', such as ligand binding, protonation, or interactions with other proteins (Dokholyan, 2016; Kern and Zuiderweg, 2003; Schueler-Furman and Wodak, 2016). The main parameters that determine the allosteric behavior are thus, 1) the relative stabilities (or populations) of all the states accessible to the system including those corresponding to active and inactive conformations (with respect to ligand binding for instance), 2) the time scales and energy barriers associated with the transitions between states, and 3) the binding affinities of the ligands/effectors or conditions, which may modify the set of dominant states, and thereby remodel the energy landscape of the system (Hilser et al., 2012; Motlagh et al., 2014). However, much remains unknown about these

important parameters. What is the role of thermodynamics, e.g. stabilizing/destabilizing different states of the system, versus the role of kinetics e.g. the time scales and energy barriers associated with the transitions between states? What are the relative contributions of entropy and enthalpy to the allosteric free energy? Is there a special role in allostery for protein intrinsic disorder? Are molecular machines a distinct category of allosteric systems? What can we learn about allostery from simple toy materials? These are some of the questions that the workshop set out to scrutinize.

Bolhuis and Faccioli reported progress in simulation algorithms for investigating and sampling rare events such as those associated with protein folding or unfolding, or with conformational transitions between active and inactive states in some allosteric systems. Such events may involve high free-energy barriers and long transition times that are not accessible by classical molecular dynamics (MD) simulations, even with the help of advanced high-performance computers, and therefore require the use of specialized sampling techniques involving various levels of approximations (Amaro et al., 2007; Markwick and McCammon, 2011; Pontiggia et al., 2015; Proctor et al., 2015).

The advantage of the enhanced sampling algorithms developed in the **Bolhuis** group is that they require no prior knowledge of the reaction coordinates (the main geometric parameters that change during the reaction process), which is usually not available. Using only information on the initial and final states, these algorithms generate the collection of trajectories that connect these two states, and employ the transition path sampling (TPS) algorithm (Bolhuis et al., 2002), which incorporates methods for selecting efficient moves along the energy landscape (Brotzakis and Bolhuis, 2016), to sample the shortest transition

paths across these trajectories. Those can then be scrutinized for pertinent reaction coordinates, and used to estimate the transition rates by evaluating the trajectory fluxes (Moroni et al., 2005). An example of the application of TPS to sample the light-induced conformational transition of the Photoactive Yellow Protein (PYP) (Vreede et al., 2010), a water-soluble blue-light photoreceptor from *H. halophila*, is illustrated in **Figure 1A**. All the path-finding methods developed by these authors are available in the OpenPathSampling software (Swenson et al., 2018).

The self-consistent path sampling (SCPS) method of **Faccioli** and collaborators affords further reductions in computational cost, but at the price of additional approximations, making it possible to simulate very slow conformational transitions of very large protein systems, using state-of-the-art atom-based force fields. Their method is based on a set of self-consistent stochastic equations of motion from which reaction pathways are generated by an iterative procedure (Orioli et al., 2017). The method also outputs a stochastic estimate of the reaction coordinates, and enables estimation of the potential of mean force of arbitrary collective coordinates. A variant of the SCPS methods was used to characterize the extremely slow conformational transition of the ~400 residues alpha1-antitrypsin of the serpin family (Cazzolli et al., 2014).

The two sampling methods, originally developed to model protein folding/unfolding reactions, represent important advances. But their potential to yield reliable mechanistic descriptions of the conformational transitions of allosteric systems still needs confirmation. Such confirmations could be obtained by applying the simulation procedures to systems for which the allosteric transition has been characterized experimentally, thereby enabling

direct comparison with the results of the simulations.

Stock and colleagues have done precisely that. Recent time-resolved infrared spectroscopy experiments on a photo switchable PDZ2 domain have indicated that the allosteric transition in this system occurs on multiple timescales (Buchli et al., 2013). Moreover, NMR relaxation experiments on the closely related PDZ3 domain revealed allosteric couplings between the binding pocket and the C-terminus (Petit et al., 2009). To gain insight into the underlying process, Stock employed exhaustive non-equilibrium molecular dynamics simulations to derive a time-dependent description of this transition (Buchenberg et al., 2017; Stock and Hamm, 2018). Results revealed that the structural and dynamic changes undergone by the system are highly non-linear and occur in a non-local fashion, in excellent agreement with the experimental data. This in turn led the authors to propose similarities with the process of downhill protein folding and to question the soundness of interpreting allosteric transitions in terms of well-defined pathways for propagating the conformational changes, as commonly done in the literature.

The experimental and modeling work by **Hilser** and colleagues on proteins with intrinsically disordered (ID) regions also undermines a strict pathway interpretation of allosteric transitions. Intrinsically disordered proteins represent a functional oddity because they lack stable tertiary structures, but represent nevertheless allosteric systems that play a central role in signaling processes (Ferreon et al., 2013; Garcia-Pino et al., 2010; Lum et al., 2012; Motlagh et al., 2014; Sevcsik et al., 2011). Investigating the mechanism of transcriptional regulation of the glucocorticoid receptor (GR), a protein involved in signaling whose functionally important N-terminal domain (NTD) is intrinsically disordered, the authors

showed that this protein is able to allosterically regulate function by simultaneously tuning transcriptional activation and repression (Li et al., 2017). This allosteric regulation is achieved by producing translational isoforms differing only in the length of the disordered domain and displaying different DNA-binding affinities and transcriptional activities that are uncorrelated to each other. Based on biophysical measurements analyzed in the framework of the ensemble model of allostery championed by the authors (Motlagh et al., 2014), compelling evidence was presented that this uncorrelated behavior is enabled through a mechanism of 'energetic frustration', whereby opposing energetic couplings between the structured domains and the disordered regions compete to modulate the overall response, as illustrated in **Figure 1B**.

Bahar reviewed approaches based on elastic network models (ENMs), which have demonstrated the significance of soft collective modes of motion in enabling allosteric regulation of protein systems (Bahar et al., 2007; Bahar et al., 2017). These approaches are in line with the ensemble view of allostery, but focus on conformational ensembles sampled by thermal fluctuations near the native state minimum of the energy landscape. The motions described by such ensembles can be evaluated by normal mode analysis (NMA) at full atomic detail (Go et al., 1983). However, the coarse-graining of the energy landscape with the help of ENMs permits sampling a relatively broad subspace of conformers and yields a unique analytical solution for the spectrum of modes for a given protein fold. The modes at the low frequency end of the spectrum (soft modes) are particularly relevant to allostery, as they are both highly cooperative and robustly defined by the overall architecture of the system.

Applying ENMs to several systems showed that the conformational changes of proteins elicited by ligand binding closely overlap with one or more of the soft modes accessible in the unbound form (Bahar et al., 2010; Tobi and Bahar, 2005), as illustrated for adenylate kinase (Temiz et al., 2004) (**Figure 1C).** The soft modes have therefore been described as 'paths' in conformational space enabling the allosteric transitions (Meireles et al., 2011), suggesting in turn that the ability to favour such soft modes may have played a role in the evolutionary selection of modules and domains that lend themselves to allosteric regulation.

McLeish presented work focusing on allosteric control enabled solely through the modulation of thermal fluctuations and the resulting entropy changes, induced by ligand binding. A feature of this mechanism, first formalized by Cooper & Dryden (1984) (Cooper and Dryden, 1984) and termed *thermal fluctuations allo*stery by the author or '*dynamic allostery'* elsewhere (Guo and Zhou, 2016; Kern and Zuiderweg, 2003; Schueler-Furman and Wodak, 2016), is that soft global modes of motion rather than more local ones are recruited to enable allosteric cooperativity. As seen above, such soft modes may be readily described by coarse-grained models like those of Bahar and co-workers (Bahar et al., 2010) and others (Hawkins and McLeish, 2004; Zhu et al., 2011). To further investigate the implications of 'fluctuation allostery' McLeish uses the coarsest possible toy model of a protein, consisting of just one (harmonic) internal degree of freedom. This simple unit, termed *allosteron,* of which a real example was described in the eighties (Onan et al., 1983), features one or more ligand binding sites and can also oligomerize. Crucially, it undergoes internal fluctuations modified by the binding of each ligand (**Figure 2A**). The author demonstrates that using the classical approximation to the harmonic-oscillator

partition function yields reasonable estimates of the allosteric free energy between two ligands bound to such a system, which contain no enthalpic terms. Extensions of the allosteron model have been helpful in identifying the physical origin of associated phenomena, such as the coupling of global and local vibrational modes in dynamic allostery of proteins (Hawkins and McLeish, 2006), the negative cooperativity of the CAP (Catabolite Activator Protein) homodimer (Toncrova and McLeish, 2010), or the sequence of effector-binding events in allosteric multi-protein assemblies (McLeish et al., 2018).

The important role of protein dynamics in enabling allosteric regulation was further highlighted by the computational studies of **Palermo and McCammon,** performed on the large multi-domain CRISPR-Cas9 system (Palermo et al., 2016, 2017a; Palermo et al., 2017b), the centerpiece of a recently emerged transformative genome editing technology (Chen and Doudna, 2017). In this multi-domain system, the endonuclease Cas9 associates with single-guide RNAs to site-specifically recognize and cleave any DNA sequence bearing a Protospacer Adjacent Motif (PAM) sequence. RNA-mediated binding to this sequence initiates DNA association and cleavage, with the latter performed by two spatially distant domains of the protein, HNH and RuvC, via a concerted mechanism. From MD trajectories of the CRISPR-Cas9 complex bound to PAM and for its analogue crystallized without PAM (Palermo et al., 2017b), the authors computed the generalized correlations (GC), capturing both linear and non-linear correlated motions of the system. Using the GC coefficients as edge weights, a residue dynamic network was built from each trajectory. Analyzing these networks revealed tighter communication (and increased correlated motions) between the HNH and RuvC domains in the presence of PAM (**Figure 1E**). This led the authors to conclude that PAM binding to CRISPR-Cas9 plays a key role in

triggering the interdependent conformational dynamics of HNH and RuvC, likely enabling the concerted cleavage of the DNA strands(Palermo et al., 2018). It furthermore allowed the identification of residues responsible for the information relay. Mutating two of these residues (K775A and R905A) was shown to decrease off-target cleavage of partially complementary DNAs (Chen and Doudna, 2017), opening an avenue for modulating the activity of CRISPR-Cas9 systems.

In their contribution to this workshop review, **Nussinov and colleagues** adhere to the ensemble model of allostery, and view allosteric regulation as resulting from perturbations of the inactive (or active) conformational ensembles leading to activation (or inactivation) via a 'population shift' (Gunasekaran et al., 2004; Tsai and Nussinov, 2014). They also acknowledge the role of dynamics, but argue against the concept of *dynamic allostery* discussed above, which involves no changes between distinct conformational states (see also (Kern and Zuiderweg, 2003)). In Nussinov's view, only distinct states, corresponding to local minima of the native free energy landscape, can contribute to functional allostery, because specific functions are performed by distinct protein conformations (Nussinov and Tsai, 2015) as exemplified in the supplementary **Figure S1.** In support of their view they enumerate reasons for failing to observe conformational changes in some prominent allosteric systems. These reasons include crystal-packing effects, non-native crystallization conditions, which may stabilize the inactive state or destabilize the effector-bound active conformation and therefore trap a state exhibiting no conformational change. Also mentioned are, inadequate accounting for disordered regions, ignoring synergistic effects between allosteric effectors, and too short molecular dynamics simulations.

Novel mechanistic insights into the allosteric transitions of large multi-subunit molecular machines were derived from the experimental work of **Horovitz.** The efficiency of molecular machines is path-dependent. Understanding how these machines work therefore requires characterizing the intermediate and transition states of the allosteric switch reaction. In the case of ATP-consuming bio-molecular machines, which are often multimeric proteins, a key issue is whether they undergo concerted (Monod et al., 1965), sequential (Koshland et al., 1966), or probabilistic conformational changes. Horovitz showed how recent advances in single-molecule techniques and native mass-spectrometry finally made it possible to distinguishing between these models. Using these techniques enables quantification of the populations of co-existing states with different numbers of bound ligand molecules, giving rise to a particular degree of fractional saturation (**Figure 1D**). Given these populations, it is possible to determine the ligand binding constants for a multimeric protein and thus, to infer its allosteric mechanism (Gruber and Horovitz, 2018). Results showed that the ATP-promoted allosteric transitions of the homo-heptameric rings of GroEL are concerted (Dyachenko et al., 2013). Phi-value analysis, shown to be useful for studying protein folding reactions, revealed two parallel pathways for the allosteric transition of this protein (Gruber and Horovitz, 2016). A different approach based on an Arrhenius analysis of ATP hydrolysis by the group II chaperonin CCT/TRiC, the eukaryotic homolog of GroEL, revealed that the intra-ring conformational changes in this protein associated with ATP hydrolysis are sequential (Gruber et al., 2017). Structural features and possible evolutionary pressure that may underlie these intriguing differences between the two chaperonins were briefly discussed.

Thought-provoking investigations of the architectural principles and properties of allosteric materials were presented by **Wyart.** Considering allostery as the process whereby ligand binding at one site of a protein transmits a signal to a distant functional site, the authors investigate this process from a purely physical perspective. Among the questions that they set out to answer were how materials can be designed to carry mechanical information over long distances, or what allosteric pathways may be optimized for? The approach consists in using *in silico* 'evolution' schemes to optimize elastic toy materials, two-and three-dimensional spring networks, for carrying out a specific 'function' (Flechsig, 2017; Rocks et al., 2017; Yan et al., 2017a; Yan et al., 2017b, 2018). A surprising result from these *in silico* experiments is that the type of 'function' greatly affects the resulting architectures. Optimizing the networks for a geometric task, by selecting network structures where binding a ligand leads to a defined displacement on the other side of the network ('active site'), yields networks displaying a powerful lever at the active site, where the signal is required (Flechsig, 2017; Rocks et al., 2017; Yan et al., 2017a; Yan et al., 2017b). This lever has distinctive structural properties (between those of a solid and liquid) and may represent a potential candidate mechanism for allosteric proteins in which motion such as that for opening or closing a channel is required (**Figure 2B**). Completely different architectures evolve when the networks are optimized for cooperative binding energy between the allosteric and active site (Yan et al., 2017a). These evolved architectures feature a very soft elastic mode that extends throughout the structure. In addition, most of the response tends to be captured by a single normal mode, as observed in some allosteric proteins. Crucially, it was found that to induce cooperativity, the frequency of this mode must adopt moderate values, with the predicted optimal frequency depending on the linear size of the system. Despite the simplicity of the investigated materials, one is left with the

impression that these *in silico* evolution approaches should be very useful for formulating key questions about real allosteric systems that may be addressed experimentally.

Allostery and signalling

The allosteric behavior of proteins and protein assemblies plays a key role in signaling processes. Unraveling the mechanistic underpinning of this behavior should therefore lead to improved understanding of how signaling events are relayed and regulated, and enable their modulation with promising pharmaceutical avenues for targeting human disease (Dokholyan, 2016)

Stote and Dejaegere reported findings on the mechanism of allosteric regulation of retinoic acid receptors (RARs), members of the nuclear receptor (NR) superfamily implicated in the transcriptional cascades underlying many physiological phenomena, such as cell differentiation and growth (Brelivet et al., 2012; Helsen and Claessens, 2014). Although retinoic acid has been considered the primary regulator of RARs, phosphorylation of the ligand-binding domain (LBD) has been shown to modulate downstream nuclear signaling by phosphorylation of the regulatory N-terminal domain (NTD) (**Supplementary Figure S2**). Crystallographic studies of phospho-mimetic mutations of RARγ (S371E) and molecular dynamics simulations showed that phosphorylation of the RARγ (and RARα) receptors of this family leads to subtle changes in the dynamic properties of the protein without producing significant conformational rearrangements (Chebaro et al., 2013; Chebaro et al., 2017). It was furthermore proposed that a conserved long alpha-helix plays a key role in mediating the allosteric communication between sites in these receptors and

likely in other members of the nuclear receptor superfamily where the long helix in question is well conserved.

Cecchini & Changeux presented a strategy for modeling allosteric transitions in proteins. This strategy involves adding or removing an agonist from the binding site of an allosteric protein and using unbiased MD simulations to capture the spontaneous transition/relaxation of the system to a distinct physiological state **(Supplementary Figure S3)**. The approach was applied to the pentameric ligand-gated ion channels (pLGICs), representing typical allosteric membrane proteins that serve as signal transducers in neurotransmitter-mediated intercellular communication. In these systems, the activation/relaxation MD protocol was used to explore the pore-closing transition or *un-gating* of the prokaryotic proton-gated channel GLIC (Nury et al., 2010). Similarly, MD relaxation of the open form of the eukaryotic glutamate-gated ion channel, (GluCl) upon removal of the positive allosteric modulator ivermectin, was shown to promote partial closure of the ion pore through a complex quaternary mechanism involving global receptor *twisting* and a radial expansion (*blooming)* of the extracellular domain (Calimet et al., 2013). A more extended relaxation of the same channel in the absence of ivermectin captured the full closing motion that is consistent with the ligand-free GluCl x-ray structure (Martin et al., 2017). Using the same approach, the gating mechanism of pLGICs was explored also in the forward direction (from resting to active), revealing a correlation between orthosteric agonist binding and ion-pore opening (Yoluk et al., 2015; Yuan et al., 2016). The MD-based activation/relaxation protocol thus appears as a useful approach for exploring the allosteric transitions at atomic resolution in these large important systems,

despite its high computational costs and the fact that it collects only a limited number of transition events.

Rational design of allosteric systems and identification of allosteric sites.

Several approaches for the rational design of allosteric systems, allosteric switches and allosteric sensors, were described by Dokholyan, Berezovsky, Karanicolas, and Plaxco.

Dokholyan described new optogenetic and chemogenetic tools for controlling individual proteins and signaling cascades in living cells (Dagliyan et al., 2017; Dagliyan et al., 2013; Dagliyan et al., 2016). The approach consists of using computational procedures to identify solvent accessible allosteric sites (Proctor et al., 2015) on a target protein and physically engineering naturally occurring light-sensitive or ligand-sensitive domains into these sites. Light or a ligand are then used to modulate structural disorder in these domains, which, in turn, affects the active site of the target protein, switching it between inactive (increased disorder) and active (less disorder) states. In the illustrated examples (**Figure 3A**) the small naturally-occurring light-sensitive LOV2 domain, and the rapamycin-responsive uniRapR domain, were respectively engineered into several kinases involved in cell motility (Dagliyan et al., 2013). Light and rapamycin were then used to respectively, inactivate and activate the target proteins, with the resulting effects on cell motility directly monitored by imaging techniques.

The computational approach presented by **Berezovsky** quantifies the configurational work exerted in different parts of a protein as a result of ligand binding to a known or putative allosteric site and can be used to infer allosteric sites, ultimately enabling the design of

effector molecules (Guarnera and Berezovsky, 2016a, b). In this approach, an approximation similar to those described by Bahar and McLeish is used to model the protein native state dynamics. The protein force field is represented by a simple Cα-based harmonic potential, and the presence of a ligand at the allosteric site is modeled by locally restraining residue pairs at the binding site. Next, the dynamics of the ligand-free and ligand-bound proteins are described using normal mode analysis, from which a set of relevant normal modes is derived. These modes are then used to evaluate the so-called 'allosteric potential', defined as the mean work exerted on a residue as a result of the local motion of its neighbors. Lastly a per-residue "allosteric free energy" is computed from the difference between the ligand-free and ligand-bound conformational ensembles sampled by the relevant modes. Extension of the method to identify the effect of allosteric mutations and its application to the regulation of the activity of the insulin-degrading enzyme, (Guarnera and Berezovsky, 2016b; Kurochkin et al., 2017), were also mentioned. The extended method is implemented in the AlloSigMA, (http://allosigma.bii.a-star.edu.sg/home/) web-server (Guarnera et al., 2017), which can be used as a first approach for investigating allosteric effects on protein activity elicited by ligands or mutations, or for identifying potential new allosteric sites and candidates for allosteric mutations (Tee et al., 2018).

Karanicolas, on the other hand, described a method for building molecular switches, which involves the chemical rescue of the active conformation of a protein. In this procedure, a disruptive mutation (often of a hydrophobic residue important for protein stability) introduced into the protein is rescued by addition of a small molecule that

complements the deleted atoms. Proof-of-concept for this approach was demonstrated by introducing a (deactivating) tryptophan-to-glycine mutation into an enzyme, then showing that activity could be restored by adding indole to complement the resulting cavity (Deckert et al., 2012). The generality of this approach for building allosteric control into proteins other than enzymes was then explored by developing a cell-based reporter assay. This allowed for screening of many W→G mutations to determine which would attenuate protein activity, and then for testing which of these mutants could subsequently be rescued using indole. A suite of computational and experimental methods, collectively led to the insight that protein structure and function were most frequently modulated *indirectly* through control of protein stability (Xia et al., 2013). Addition of indole in these allosteric cases served not to revert a discrete conformational change, but rather as an allosteric ligand that rescues activity by inducing the protein to refold to its original conformation (Budiardjo et al., 2016), thereby representing an excellent illustration of the ensemble model of allostery.

Plaxco described how allostery and cooperativity may be leveraged to engineer a wide range of artificial optical, biochemical and electrochemical biosensors. Among the examples used to illustrate the approach was the rational design and engineering of a synthetic DNA-based nanodevice containing up to four interacting binding sites that can load and release a cargo over narrow concentration ranges, and whose affinity could be finely controlled via both allosteric effectors and environmental cues like pH and temperature (Mariottini et al., 2017). In another example, catalytic DNAzyme sequences (e.g. peroxidase-like DNAzymes) were combined with the consensus sequence recognized by specific transcription factors (either TATA binding protein or the microphthalmia-associated transcription factor). The resulting constructs exhibited, respectively, a more

stable catalytically inactive conformation unable to bind the cognate transcription factor, and a less stable conformation competent to bind it. The presence of the transcription factor pushes the equilibrium between these states towards the catalytically active one, in a manner that can be finely controlled further by optimizing the original design (Adornetto et al., 2015).

Kozakov presented an approach for identifying allosteric binding sites (also denoted as cryptic sites) in ligand-free protein structures, and predicting their drug binding potential**.** The method involves the identification of binding hotspots on the protein surface. These hotspots represent clusters of low energy binding poses for small organic molecular probes of various shapes, sizes and polarity, generated by their FTsite computational procedure (Ngan et al., 2012). Applying FTsite to protein structures with known allosteric sites (Cimermancic et al., 2016) it was found that the ligand-free apo structures generally feature binding hotspots for the tested small molecular probes that are in close proximity to the known allosteric sites (**Supplementary Figure S4**). Of those, the more highly populated hotspot clusters (≥16 low energy poses) were deemed druggable, e.g. can be targeted by ligands with sufficient affinity (Kozakov et al., 2015). The authors also reported that regions of protein structures close to cryptic binding sites are significantly more flexible than regions surrounding any other potential binding hotspots detected by their procedure (Beglov et al., 2018). This increased flexibility seems to be linked to missing loops or side chains of less reliably modeled regions of the corresponding X-ray structures, suggesting that such regions may be good cryptic binding site candidates.

Lastly, among the notable poster presentations, three reported analyses of the dynamics and allosteric regulations in important multi-subunit enzymes from various origins. **Rivalta** and colleagues used classical MD simulations and a community network analysis (Sethi et al., 2009), not unlike that of Palermo & McCammon, to elucidate the allosteric regulation in the imidazole glycerol phosphate synthase (IGPS) from *T. maritima* (Rivalta et al., 2012). This analysis stimulated single site mutagenesis experiments and allosteric inhibitor design (Rivalta et al., 2012) (**Supplementary Figure S5** for details). **Gkeka** and collaborators described potentially important findings from combined experimental and computational analyses on the allosteric modulation of the lipid phosphoinositide 3-kinase alpha, PI3Kα , which plays a pivotal role in cell proliferation and is a target for anti-cancer drug development (see **Supplementary Figure S6** for details]. They discovered a ligand-binding site distinct from the enzyme active site capable of inhibiting a cancer-associated PI3Kα mutant responsible for enzyme over activation. Ligand binding to this site was found to modulate the membrane binding domain of the protein, and not the active site, opening the avenue for designing selective inhibitors of protein-membrane interactions in this and other systems (Gkeka et al., 2014; Gkeka et al., 2015). **Panecka-Hofman and Wade** reported preliminary results on the dynamic allosteric coupling between distant residues of Pteridine reductase 1 (PTR1), a folate pathway enzyme unique to trypanosomatid parasites (Panecka-Hofman et al., 2017) (**Supplementary Figure S7**). The 4th poster reported progress towards gaining insight into the allosteric regulation of taste GPCRs (Di Pizio et al., 2016) (**Supplementary Figure S8**).

Concluding remarks

In this collection of contributions presented at the CECAM workshop, we endeavored to provide an overview of the current understanding of allosteric processes and its perceived limitations. We also described how this still incomplete understanding is exploited more or less successfully to illuminate the underlying molecular mechanisms, explain cellular processes, design molecular sensors and inform drug design efforts.

The concept of allostery has evolved significantly since the first allosteric proteins have been characterized (Motlagh et al., 2014; Schueler-Furman and Wodak, 2016). We now have a better grasp of the important functional role of protein dynamics and, in particular, the role of protein intrinsic disorder. We also have more powerful computational and experimental tools for sampling significantly populated states of complex protein systems.

Notwithstanding these advances, current computational methods are still unable to chart the free energy landscape of allosteric systems in an unbiased way, e.g. without prior knowledge of some significantly populated states of the system. Even when such knowledge is available, computational procedures employ various levels of approximations to sample the conformational transition paths between these states, as illustrated by the contributions of Bulhuis, Faccioli, and Stock, and work of groups employing Markov State Models (MSM) (Chodera and Noe, 2014; Pande et al., 2010). Specific approaches depend moreover on the size and complexity of the systems under study, making it difficult to evaluate the information they provide about the identified transition paths. To enable such evaluation it would be useful to come up with a few allosteric protein systems with well-characterized active and inactive states, to which different computational methods for

charting the allosteric transition paths could be applied, results compared, and eventually evaluated against experimental data.

Particularly useful would be data derived from phi-value type analyses. Such analyses measure the changes in the activation energy of unfolding and the free energy of unfolding brought about by mutations, and those are used to characterize the transition states and intermediates of protein folding reactions (Fersht et al., 1992). Employing similar analyses to characterize the transition state of an allosteric pathway was suggested during the meeting, but not further elaborated on, although there have indeed been insightful precedents. Eaton et al. (Eaton et al., 1991) were the first to apply such analyses, generally referred to as rate-equilibrium linear free-energy relationships (LFERs), to allostery. Using pH and ligand states instead of mutations to perturb the kinetics and thermodynamics of the allosteric transition in hemoglobin, they showed that the transition state of the R<->T quaternary conformational change had closer thermodynamic properties to those of the R than the T conformations, validating a much earlier computational study, based on a crude analysis of the surface areas buried between the subunits (Janin and Wodak, 1985). A subsequent study of Yifrach and Horovitz (Yifrach and Horovitz, 1998) employed a genuine phi-value analysis, involving a limited number of mutations, to map the transition state of the allosteric pathway of GroEL. LFERs derived from perturbations, notably by a series of site-specific mutations, were used to map or the transition state of the gating reaction pathway of the muscle acetylcholine receptor (Grosman et al., 2000), yielding detailed information on the gating mechanism, described as involving a wave-like conformational change.

Computational approaches to the seemingly more tractable problem of identifying paths that mediate allosteric 'communication' between sites in a protein would also benefit from a more objective benchmarking. Although fundamentally different from allosteric transition paths on the free energy landscape, identifying communication paths also involves sampling the free energy landscape, but only in the vicinity of the stable 'end' states, and then quantifying the correlated motions of the corresponding conformational ensembles. But here too, computational procedures and the set of investigated systems tend to differ significantly between authors. Assessing the agreement between communication paths identified by different methods in the same set of allosteric systems should therefore be very informative. Since even in a highly structured protein 'communication' between sites is likely mediated by multiple paths (Guo et al., 2015; Taylor et al., 2016), the questions of whether a given path can be rigorously validated against experimental data, or whether its specific role can be rationalized, need to be critically evaluated.

Two distinct but complementary conceptual frameworks for probing the mechanism of allosteric regulation, highlighted in this review, deserve special mention. One considers allosteric regulation as enabled by the so-called 'soft' modes of collective motions sampled by thermal fluctuations near the native state minimum, usually of highly structured protein systems. These soft modes are estimated computationally from experimental structures, using coarse-grained elastic network models (ENM), which strongly depend on the reference structure (usually the experimentally determined one). With skeptics, wary of such coarse-grained models, one would argue that ENM and the underlying conceptual framework have been quite instrumental not only in capturing the conformational transitions associated with the allosteric regulation of complex protein systems, but also in modeling

the entropic contributions to the allosteric free energy, and potentially for predicting allosteric binding sites in protein systems, as reported by several contributions to this review and references therein.

The other conceptual framework refers to the so-called ensemble model of allostery, which focuses entirely on the thermodynamic analysis of the energy landscape of allosteric systems, including those featuring intrinsic disorder. It is thus of very broad applicability. As already mentioned, the main task of such analysis is quantifying the relative populations (stabilities) of all the states accessible to the system and how this population landscape is modified by ligand/effector binding, or disorder-to order transitions. Focusing on these thermodynamic properties is amply justified. In many systems, the rate limiting step of the allosteric transition elicited by effector binding, may indeed be governed by the concentration (population) of the pre-existing ligand-binding competent state of a protein, rather than by the free energy barrier of the conformational transition it needs to undergo to adopt this state. As illustrated here by a number of contributions, fine-tuning the relative populations of the active and inactive states of protein or nucleic acid systems and the binding affinities of allosteric effectors, are very effective ways, by nature or in the laboratory, to design systems undergoing allosteric regulation of different levels of complexity and versatility.

Clearly, allosteric regulation still needs to deliver many of its secrets. An advantage of allosteric regulation over regulation involving gene expression is its shorter response time to changing conditions. One may therefore wonder if this may determine the set of properties of allosteric systems, such as the existence of soft collective motions or

population levels of relevant states, that evolution tends to select. Are all proteins allosteric, as some have suggested (Gunasekaran et al., 2004)? Are molecular machines a special category of allosteric systems? And lastly, how much can we learn about the very fundamental requirements of allostery from simple toy materials? These are only some of the many intriguing questions to address, going forward.

Acknowledgements

The Centre Européen pour le Calcul Atomic et Moléculaire (CECAM), and the Swiss Industry Science Fund (formerly the KGF) are gratefully acknowledge for support of the CECAM workshop on Computational Approaches to Investigating Allostery.

MC and JPC acknowledge funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No. 720270 (HBP SGA1).
NVD acknowledge support by the National Institutes of Health Grants R01GM114015, R01GM064803, and R01GM123247, and IB, acknowledges support from grants P41GM103712 and P30DA035778. **RN, CJT**, and **HJ** acknowledge support in whole or in part from Federal funds from the National Cancer Institute, National Institutes of Health, under contract number HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This research was supported in part by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research. **GP** and **JAM**, thank Dr. Clarisse G. Ricci, Prof. Victor S. Batista, Dr. Amendra Fernando, Prof. Martin Jinek and Rajshekhar Basak for having contributed to their study, and the NIH (grant: NIG GM31749) for support. **ADP**, and **MYN** acknowledge funding from ISF 494/16 and ISF-NSFC 2463/16 grants. **PG** acknowledges funding from the NSRF 2007–2013, the European Regional Development Fund and national resources, under the grant "Cooperation" [No. 09ΣΥΝ 11-675], the Marie Curie Reintegration Grant (FP7-PEOPLE-2009-RG, No 256533). An AACR Judah Folkman Fellowship for Cancer Research in Angiogenesis (08-40-18-COUR) is gratefully acknowledged. We thank PRACE for awarding us access to the computational facility CURIE at GENCI@CEA, France are grateful for support form the LinkSCEEM-2 project, funded by the European Commission under the 7th Framework Programme (INFRA-2010-1.2.3) Virtual Research Communities, and Combination of Collaborative Project and Coordination and Support Actions (CP-CSA) under grant agreement no RI-261600. **JPH**, and **RCW** acknowledge support from the

Polish National Science Centre (grant no. 2016/21/D/NZ1/02806) and the BIOMS program at the Interdisciplinary Center for Scientific Computing IWR, University of Heidelberg. Support of the Klaus Tschira Foundation is also gratefully acknowledged.

Author contributions

SJW wrote the manuscript on the basis of extended abstracts contributed by all authors. EP NVD and AH, helped proofreading and editing. NVD, INB, AH, JL, VJH, IB, JK, GS, PH, RHS, JE, YC, AD, MC, SO, RR, LY, CB, MW, PG, IE, GP, JAM, JPH, RCW, ADP, MYN, RN, CJT, HJ, DP, DK, TML, contributed extended abstract and Figures.

Declaration of Interest

The authors declare no competing interests.
JPH declares the following additional affiliations: Centre of New Technologies, University of Warsaw, Poland; Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies (HITS), Germany; Interdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Germany
GP declares the following additional affiliation: Department of Bioengineering, University of California Riverside, CA 92507, USA

Figure captions

Figure 1: Mechanistic underpinning of allostery: insights from computations and experiments

(A) Artist rendering of the conformational transitions network of the Photoactive Yellow Protein (PYP), the 125-residue water-soluble blue-light photoreceptor from *H. halophile*, mapped onto the energy landscape of the system using the simulation procedures of Bolhuis and collaborators**.**

(B) Frustration-based allostery in the human glucocorticoid receptor (GR), an intrinsically disordered transcription factor analysed by Li and Hilser (Li et al., 2017).
(I) Domain organization of the constitutively active GR constructs for translational isoforms,

wherein the intrinsically disordered N terminal domain (NTD) varies in length. Residues 1-97 (red) are labeled R (for Regulatory) and residues 98-420 (grey) are labeled F (for Functional). Also labeled are residues corresponding to the activation function 1 core (AF1 core) region, which is required for transcriptional activity. (**II)** Competing thermodynamic coupling in GR produces frustration. Schematic view of the thermodynamic configuration of GR. According to the displayed convention, the positive (+) signs between the DBD and F-domain, and the DBD and R-domain signify they are positively coupled; stabilization of one domain stabilizes the other. The negative (-) sign between the R- and the F- domains indicate they are negatively coupled; stabilization of one domain destabilizes the other.

(C) The closed conformation adenylate kinase observed upon ligand binding is sampled by the open form apo-structure, illustrating the work of Bahar**. (I)** Two experimentally resolved structures, unbound (left) and ligand-bound (right). **(II)** Conformer predicted by ENM (Elastic Network Model) analysis to be accessible via a soft mode to the unbound structure. Blue and green refer to different domains. The substrate is shown in orange spheres. (Adapted from (Temiz et al., 2004)).
(D) Distributions of GroEL molecules with different numbers of bound ATP molecules at different ATP concentrations from the work of Horovitz and co-workers.

(E) Allosteric regulation in CRISPR-Cas9, by Palermo and McCammon. **(I)** Dynamical network model of CRISPR-Cas9, identifying groups (or "communities") of closely correlated residues and the strength of correlation between them before (top) and upon (bottom) PAM binding. **(II)**. The allosteric path between the spatially distance HNH and RuvC domains of the Cas9 protein.

Figure 2. Allosteric toy models and allosteric materials
(A): Schematics of the allosteron model of McLeish, in binding (**a**) and self-assembly (**b**) illustrating local changes to spring constants κ, and the introduction of coupling springs between allosteron units κ_c.

(B) Illustration of the work of Wyart and collaborators: (**I)** Response (black) arrows to a stimulus (purple arrows) in random spring network decays rapidly with distance, i.e. there is little action at a distance. **(II)** Networks can be evolved in which there is specific action at a distance. Note that the response is amplified near the active site (blue arrows), indicating the presence of a lever in the structure. **(III)** Example of hinge architecture obtained while optimizing cooperativity, in which two parts of the material rotate around a hinge located at the center of the system. **(IV-VI)** Illustration of the cooperative architectures found: hinge (clothespin), shear (mint box) and twist (Rubik's cube).

Figure 3. Rational design of allosteric systems and identification of allosteric sites.
(A) Schematic diagrams illustrating the work of Doholyan and colleagues on optogenetic and chemogenetic control of target proteins using allostery and protein order-disorder transition, reprinted from (Dagliyan et al., 2016).

(B) Illustration of the approach by involving the chemical rescue of the active conformation of a protein. The example shows how mutation of a buried tryptophan to glycine leads to a structural disruption – either through a discrete conformational change or through loss of protein stability – that leads to loss of protein function. Adding exogenous indole can then complement the cavity caused by the deleted sidechain, restoring the original protein conformation and, thus, its function.

(C) Principle of the rational design and engineering of a synthetic DNA-based nanodevice described by Plaxco. Top: the designed cooperative DNA-nanodevice comprises the recognition element consisting of a triplex forming DNA sequence, which behaves like a “clamp” that binds a specific 9-base DNA ligand via the formation of both Watson-Crick and Hoogsteen base-pair interactions. The cooperative DNA-nanodevice is obtained by joining together two sequential copies of one half of such recognition element linked via a flexible 22-base, single-stranded loop (grey portion) to two sequential copies of its other half. Binding of the ligand to the first receptor decreases the entropic cost associated with the binding to the second receptor (and thus improves its affinity for the ligand). As a result,

this nanodevice shows a Hill-type cooperative response, with a Hill coefficient n_H = 2.1±0.1. Figure reproduced from (Mariottini et al., 2017).

(D) Binding hotspots of small chemical probes to flexible regions of the protein tend to correspond to cryptic binding sites**.** Example from the work of Kozakov**,** showing the mapping of hotspots identified by FTsite in the unbound structure of the catalytic subunit of the cAMP dependent protein kinase PKA (PDB ID code 2GFC, chain A) displayed in tan. Three hot spots, obtained after domain splitting, are shown as clusters of molecular probes: a cluster of 18 probes (cyan); cluster of 16 probes (magenta); cluster of 13 probes (gray). An inhibitor (yellow) is superimposed for reference.

References

Adornetto, G., Porchetta, A., Palleschi, G., Plaxco, K.W., and Ricci, F. (2015). A general approach to the design of allosteric, transcription factor-regulated DNAzymes. Chemical science *6*, 3692-3696.

Amaro, R.E., Sethi, A., Myers, R.S., Davisson, V.J., and Luthey-Schulten, Z.A. (2007). A network of conserved interactions regulates the allosteric signal in a glutamine amidotransferase. Biochemistry *46*, 2156-2173.

Bahar, I., Chennubhotla, C., and Tobi, D. (2007). Intrinsic dynamics of enzymes in the unbound state and relation to allosteric regulation. Current opinion in structural biology *17*, 633-640.

Bahar, I., Jernigan, R.L., and Dill, K. (2017). Protein Actions: Principles & Modeling. (New York, NY Taylor & Francis Group).

Bahar, I., Lezon, T.R., Yang, L.W., and Eyal, E. (2010). Global dynamics of proteins: bridging between structure and function. Annual review of biophysics *39*, 23-42.

Beglov, D., Hall, D.R., Wakefield, A.E., Luo, L., Allen, K.N., Kozakov, D., Whitty, A., and Vajda, S. (2018). Exploring the structural origins of cryptic sites on proteins. Proceedings of the National Academy of Sciences of the United States of America *115*, E3416-E3425.

Bolhuis, P.G., Chandler, D., Dellago, C., and Geissler, P.L. (2002). Transition path sampling: throwing ropes over rough mountain passes, in the dark. Annual review of physical chemistry *53*, 291-318.

Brelivet, Y., Rochel, N., and Moras, D. (2012). Structural analysis of nuclear receptors: from isolated domains to integral proteins. Molecular and cellular endocrinology *348*, 466-473.

Brotzakis, Z.F., and Bolhuis, P.G. (2016). A one-way shooting algorithm for transition path sampling of asymmetric barriers. The Journal of chemical physics *145*, 164112.

Buchenberg, S., Sittel, F., and Stock, G. (2017). Time-resolved observation of protein allosteric communication. Proceedings of the National Academy of Sciences of the United States of America *114*, E6804-E6811.

Buchli, B., Waldauer, S.A., Walser, R., Donten, M.L., Pfister, R., Blochliger, N., Steiner, S., Caflisch, A., Zerbe, O., and Hamm, P. (2013). Kinetic response of a photoperturbed allosteric protein.

Proceedings of the National Academy of Sciences of the United States of America *110*, 11725-11730.
Budiardjo, S.J., Licknack, T.J., Cory, M.B., Kapros, D., Roy, A., Lovell, S., Douglas, J., and Karanicolas, J. (2016). Full and Partial Agonism of a Designed Enzyme Switch. ACS synthetic biology *5*, 1475-1484.
Calimet, N., Simoes, M., Changeux, J.P., Karplus, M., Taly, A., and Cecchini, M. (2013). A gating mechanism of pentameric ligand-gated ion channels. Proceedings of the National Academy of Sciences of the United States of America *110*, E3987-3996.
Cazzolli, G., Wang, F., a Beccara, S., Gershenson, A., Faccioli, P., and Wintrode, P.L. (2014). Serpin latency transition at atomic resolution. Proceedings of the National Academy of Sciences of the United States of America *111*, 15414-15419.
Changeux, J.P. (1961). The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine. Cold Spring Harbor symposia on quantitative biology *26*, 313-318.
Changeux, J.P. (2011). 50th anniversary of the word "allosteric". Protein science : a publication of the Protein Society *20*, 1119-1124.
Changeux, J.P. (2012). Allostery and the Monod-Wyman-Changeux model after 50 years. Annual review of biophysics *41*, 103-133.
Changeux, J.P., and Edelstein, S.J. (2005). Allosteric mechanisms of signal transduction. Science *308*, 1424-1428.
Chebaro, Y., Amal, I., Rochel, N., Rochette-Egly, C., Stote, R.H., and Dejaegere, A. (2013). Phosphorylation of the retinoic acid receptor alpha induces a mechanical allosteric regulation and changes in internal dynamics. PLoS computational biology *9*, e1003012.
Chebaro, Y., Sirigu, S., Amal, I., Lutzing, R., Stote, R.H., Rochette-Egly, C., Rochel, N., and Dejaegere, A. (2017). Allosteric Regulation in the Ligand Binding Domain of Retinoic Acid Receptorgamma. PloS one *12*, e0171043.
Chen, J.S., and Doudna, J.A. (2017). The chemistry of Cas9 and its CRISPR colleagues. Nature Reviews Chemistry
1, 1-15.
Chodera, J.D., and Noe, F. (2014). Markov state models of biomolecular conformational dynamics. Current opinion in structural biology *25*, 135-144.
Cimermancic, P., Weinkam, P., Rettenmaier, T.J., Bichmann, L., Keedy, D.A., Woldeyes, R.A., Schneidman-Duhovny, D., Demerdash, O.N., Mitchell, J.C., Wells, J.A.*, et al.* (2016). CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. Journal of molecular biology *428*, 709-719.
Clarke, D., Sethi, A., Li, S., Kumar, S., Chang, R.W., Chen, J., and Gerstein, M. (2016). Identifying Allosteric Hotspots with Dynamics: Application to Inter- and Intra-species Conservation. Structure *24*, 826-837.
Cooper, A. (1984). Protein fluctuations and the thermodynamic uncertainty principle. Progress in biophysics and molecular biology *44*, 181-214.
Cooper, A., and Dryden, D.T. (1984). Allostery without conformational change. A plausible model. European biophysics journal : EBJ *11*, 103-109.
Dagliyan, O., Karginov, A.V., Yagishita, S., Gale, M.E., Wang, H., DerMardirossian, C., Wells, C.M., Dokholyan, N.V., Kasai, H., and Hahn, K.M. (2017). Engineering Pak1 Allosteric Switches. ACS synthetic biology *6*, 1257-1262.
Dagliyan, O., Shirvanyants, D., Karginov, A.V., Ding, F., Fee, L., Chandrasekaran, S.N., Freisinger, C.M., Smolen, G.A., Huttenlocher, A., Hahn, K.M.*, et al.* (2013). Rational design of a ligand-

controlled protein conformational switch. Proceedings of the National Academy of Sciences of the United States of America *110*, 6800-6804.
Dagliyan, O., Tarnawski, M., Chu, P.H., Shirvanyants, D., Schlichting, I., Dokholyan, N.V., and Hahn, K.M. (2016). Engineering extrinsic disorder to control protein activity in living cells. Science *354*, 1441-1444.
Deckert, K., Budiardjo, S.J., Brunner, L.C., Lovell, S., and Karanicolas, J. (2012). Designing allosteric control into enzymes by chemical rescue of structure. Journal of the American Chemical Society *134*, 10055-10060.
Di Pizio, A., Levit, A., Slutzki, M., Behrens, M., Karaman, R., and Niv, M.Y. (2016). Comparing Class A GPCRs to bitter taste receptors: Structural motifs, ligand interactions and agonist-to-antagonist ratios. Methods in cell biology *132*, 401-427.
Dima, R.I., and Thirumalai, D. (2006). Determination of network of residues that regulate allostery in protein families using sequence analysis. Protein science : a publication of the Protein Society *15*, 258-268.
Dokholyan, N.V. (2016). Controlling Allosteric Networks in Proteins. Chemical reviews.
Dyachenko, A., Gruber, R., Shimon, L., Horovitz, A., and Sharon, M. (2013). Allosteric mechanisms can be distinguished using structural mass spectrometry. Proceedings of the National Academy of Sciences of the United States of America *110*, 7235-7239.
Eaton, W.A., Henry, E.R., and Hofrichter, J. (1991). Application of linear free energy relations to protein conformational changes: the quaternary structural change of hemoglobin. Proceedings of the National Academy of Sciences of the United States of America *88*, 4472-4475.
Falke, J.J., and Piasta, K.N. (2014). Architecture and signal transduction mechanism of the bacterial chemosensory array: progress, controversies, and challenges. Current opinion in structural biology *29*, 85-94.
Ferreon, A.C., Ferreon, J.C., Wright, P.E., and Deniz, A.A. (2013). Modulation of allostery by protein intrinsic disorder. Nature *498*, 390-394.
Fersht, A.R., Matouschek, A., and Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. Journal of molecular biology *224*, 771-782.
Flechsig, H. (2017). Design of Elastic Networks with Evolutionary Optimized Long-Range Communication as Mechanical Models of Allosteric Proteins. Biophysical journal *113*, 558-571.
Frauenfelder, H., Parak, F., and Young, R.D. (1988). Conformational substates in proteins. Annual review of biophysics and biophysical chemistry *17*, 451-479.
Garcia-Pino, A., Balasubramanian, S., Wyns, L., Gazit, E., De Greve, H., Magnuson, R.D., Charlier, D., van Nuland, N.A., and Loris, R. (2010). Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. Cell *142*, 101-111.
Gkeka, P., Evangelidis, T., Pavlaki, M., Lazani, V., Christoforidis, S., Agianian, B., and Cournia, Z. (2014). Investigating the structure and dynamics of the PIK3CA wild-type and H1047R oncogenic mutant. PLoS computational biology *10*, e1003895.
Gkeka, P., Papafotika, A., Christoforidis, S., and Cournia, Z. (2015). Exploring a non-ATP pocket for potential allosteric modulation of PI3Kalpha. The journal of physical chemistry. B *119*, 1002-1016.
Go, N., Noguti, T., and Nishikawa, T. (1983). Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proceedings of the National Academy of Sciences of the United States of America *80*, 3696-3700.

Grosman, C., Zhou, M., and Auerbach, A. (2000). Mapping the conformational wave of acetylcholine receptor channel gating. Nature *403*, 773-776.
Gruber, R., and Horovitz, A. (2016). Allosteric Mechanisms in Chaperonin Machines. Chemical reviews *116*, 6588-6606.
Gruber, R., and Horovitz, A. (2018). Unpicking allosteric mechanisms of homo-oligomeric proteins by determining their successive ligand binding constants. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *373*.
Gruber, R., Levitt, M., and Horovitz, A. (2017). Sequential allosteric mechanism of ATP hydrolysis by the CCT/TRiC chaperone is revealed through Arrhenius analysis. Proceedings of the National Academy of Sciences of the United States of America *114*, 5189-5194.
Guarnera, E., and Berezovsky, I.N. (2016a). Allosteric sites: remote control in regulation of protein activity. Current opinion in structural biology *37*, 1-8.
Guarnera, E., and Berezovsky, I.N. (2016b). Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. PLoS computational biology *12*, e1004678.
Guarnera, E., Tan, Z.W., Zheng, Z., and Berezovsky, I.N. (2017). AlloSigMA: allosteric signaling and mutation analysis server. Bioinformatics *33*, 3996-3998.
Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? Proteins *57*, 433-443.
Guo, J., Pang, X., and Zhou, H.X. (2015). Two pathways mediate interdomain allosteric regulation in pin1. Structure *23*, 237-247.
Guo, J., and Zhou, H.X. (2016). Protein Allostery and Conformational Dynamics. Chemical reviews.
Hawkins, R.J., and McLeish, T.C. (2004). Coarse-grained model of entropic allostery. Physical review letters *93*, 098104.
Hawkins, R.J., and McLeish, T.C. (2006). Coupling of global and local vibrational modes in dynamic allostery of proteins. Biophysical journal *91*, 2055-2062.
Hayouka, Z., Rosenbluh, J., Levin, A., Loya, S., Lebendiker, M., Veprintsev, D., Kotler, M., Hizi, A., Loyter, A., and Friedler, A. (2007). Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium. Proceedings of the National Academy of Sciences of the United States of America *104*, 8316-8321.
Helsen, C., and Claessens, F. (2014). Looking at nuclear receptors from a new angle. Molecular and cellular endocrinology *382*, 97-106.
Hilser, V.J., Wrabl, J.O., and Motlagh, H.N. (2012). Structural and energetic basis of allostery. Annual review of biophysics *41*, 585-609.
Janin, J., and Wodak, S.J. (1985). Reaction pathway for the quaternary structure change in hemoglobin. Biopolymers *24*, 509-526.
Kass, I., and Horovitz, A. (2002). Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. Proteins *48*, 611-617.
Kern, D., and Zuiderweg, E.R. (2003). The role of dynamics in allosteric regulation. Current opinion in structural biology *13*, 748-757.
Koshland, D.E., Jr., Nemethy, G., and Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. Biochemistry *5*, 365-385.
Kozakov, D., Hall, D.R., Napoleon, R.L., Yueh, C., Whitty, A., and Vajda, S. (2015). New Frontiers in Druggability. J Med Chem *58*, 9063-9088.

Kurochkin, I.V., Guarnera, E., Wong, J.H., Eisenhaber, F., and Berezovsky, I.N. (2017). Toward Allosterically Increased Catalytic Activity of Insulin-Degrading Enzyme against Amyloid Peptides. Biochemistry *56*, 228-239.
Li, J., White, J.T., Saavedra, H., Wrabl, J.O., Motlagh, H.N., Liu, K., Sowers, J., Schroer, T.A., Thompson, E.B., and Hilser, V.J. (2017). Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. eLife *6*.
Link, H., Christodoulou, D., and Sauer, U. (2014). Advancing metabolic models with kinetic information. Current opinion in biotechnology *29*, 8-14.
Livesay, D.R., Kreth, K.E., and Fodor, A.A. (2012). A critical evaluation of correlated mutation algorithms and coevolution within allosteric mechanisms. Methods Mol Biol *796*, 385-398.
Lockless, S.W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. Science *286*, 295-299.
Lum, J.K., Neuweiler, H., and Fersht, A.R. (2012). Long-range modulation of chain motions within the intrinsically disordered transactivation domain of tumor suppressor p53. Journal of the American Chemical Society *134*, 1617-1622.
Mariottini, D., Idili, A., Vallee-Belisle, A., Plaxco, K.W., and Ricci, F. (2017). A DNA Nanodevice That Loads and Releases a Cargo with Hemoglobin-Like Allosteric Control and Cooperativity. Nano letters *17*, 3225-3230.
Markwick, P.R., and McCammon, J.A. (2011). Studying functional dynamics in bio-molecules using accelerated molecular dynamics. Physical chemistry chemical physics : PCCP *13*, 20053-20065.
Martin, N.E., Malik, S., Calimet, N., Changeux, J.P., and Cecchini, M. (2017). Un-gating and allosteric modulation of a pentameric ligand-gated ion channel captured by molecular dynamics. PLoS computational biology *13*, e1005784.
May, L.T., Leach, K., Sexton, P.M., and Christopoulos, A. (2007). Allosteric modulation of G protein-coupled receptors. Annual review of pharmacology and toxicology *47*, 1-51.
McLeish, T., Schaefer, C., and von der Heydt, A.C. (2018). The 'allosteron' model for entropic allostery of self-assembly. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *373*.
Meireles, L., Gur, M., Bakan, A., and Bahar, I. (2011). Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. Protein science : a publication of the Protein Society *20*, 1645-1658.
Monod, J., and Jacob, F. (1961). Teleonomic mechanisms in cellular metabolism, growth, and differentiation. Cold Spring Harbor symposia on quantitative biology *26*, 389-401.
Monod, J., Wyman, J., and Changeux, J.P. (1965). On the Nature of Allosteric Transitions: A Plausible Model. Journal of molecular biology *12*, 88-118.
Moroni, D., van Erp, T.S., and Bolhuis, P.G. (2005). Simultaneous computation of free energies and kinetics of rare events. Physical review. E, Statistical, nonlinear, and soft matter physics *71*, 056709.
Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331-339.
Ngan, C.H., Hall, D.R., Zerbe, B., Grove, L.E., Kozakov, D., and Vajda, S. (2012). FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics *28*, 286-287.
Nury, H., Poitevin, F., Van Renterghem, C., Changeux, J.P., Corringer, P.J., Delarue, M., and Baaden, M. (2010). One-microsecond molecular dynamics simulation of channel gating in a

nicotinic receptor homologue. Proceedings of the National Academy of Sciences of the United States of America *107*, 6275-6280.
Nussinov, R., and Tsai, C.J. (2015). Allostery without a conformational change? Revisiting the paradigm. Current opinion in structural biology *30*, 17-24.
Nussinov, R., Tsai, C.J., and Ma, B. (2013). The underappreciated role of allostery in the cellular network. Annual review of biophysics *42*, 169-189.
Onan, K., Rebek Jr, J., Costello, T., and Marshall, L. (1983). Allosteric Effects: Structural and Thermodynamic
Origins of Binding Cooperativity in a Subunit Model. J. Am. Chem. Soc *105*, 6759-6760.
Orioli, S., S, A.B., and Faccioli, P. (2017). Self-consistent calculation of protein folding pathways. The Journal of chemical physics *147*, 064108.
Palermo, G., Chen, J.S., Ricci, C.G., Rivalta, I., Jinek, M., Batista, V.S., Doudna, J.A., and McCammon, J.A. (2018). Key role of the REC lobe during CRISPR–Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. Quarterly Reviews of Biophysics *51*, 1-11.
Palermo, G., Miao, Y., Walker, R.C., Jinek, M., and McCammon, J.A. (2016). Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. ACS central science *2*, 756-763.
Palermo, G., Miao, Y., Walker, R.C., Jinek, M., and McCammon, J.A. (2017a). CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. Proceedings of the National Academy of Sciences of the United States of America *114*, 7260-7265.
Palermo, G., Ricci, C.G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., Batista, V.S., and McCammon, J.A. (2017b). Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. Journal of the American Chemical Society *139*, 16028-16031.
Pande, V.S., Beauchamp, K., and Bowman, G.R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. Methods *52*, 99-105.
Panecka-Hofman, J., Pohner, I., Spyrakis, F., Zeppelin, T., Di Pisa, F., Dello Iacono, L., Bonucci, A., Quotadamo, A., Venturelli, A., Mangani, S.*, et al.* (2017). Comparative mapping of on-targets and off-targets for the discovery of anti-trypanosomatid folate pathway inhibitors. Biochimica et biophysica acta *1861*, 3215-3230.
Perutz, M.F. (1970). Stereochemistry of cooperative effects in haemoglobin. Nature *228*, 726-739.
Petit, C.M., Zhang, J., Sapienza, P.J., Fuentes, E.J., and Lee, A.L. (2009). Hidden dynamic allostery in a PDZ domain. Proceedings of the National Academy of Sciences of the United States of America *106*, 18249-18254.
Pontiggia, F., Pachov, D.V., Clarkson, M.W., Villali, J., Hagan, M.F., Pande, V.S., and Kern, D. (2015). Free energy landscape of activation in a signalling protein at atomic resolution. Nature communications *6*, 7284.
Proctor, E.A., Kota, P., Aleksandrov, A.A., He, L., Riordan, J.R., and Dokholyan, N.V. (2015). Rational Coupled Dynamics Network Manipulation Rescues Disease-Relevant Mutant Cystic Fibrosis Transmembrane Conductance Regulator. Chemical science *6*, 1237-1246.
Rivalta, I., Sultan, M.M., Lee, N.S., Manley, G.A., Loria, J.P., and Batista, V.S. (2012). Allosteric pathways in imidazole glycerol phosphate synthase. Proceedings of the National Academy of Sciences of the United States of America *109*, E1428-1436.
Rocks, J.W., Pashine, N., Bischofberger, I., Goodrich, C.P., Liu, A.J., and Nagel, S.R. (2017). Designing allostery-inspired response in mechanical networks. Proceedings of the National Academy of Sciences of the United States of America *114*, 2520-2525.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. Nature reviews. Molecular cell biology *14*, 630-642.
Schueler-Furman, O., and Wodak, S.J. (2016). Computational approaches to investigating allostery. Current opinion in structural biology *41*, 159-171.
Sethi, A., Eargle, J., Black, A.A., and Luthey-Schulten, Z. (2009). Dynamical networks in tRNA:protein complexes. Proceedings of the National Academy of Sciences of the United States of America *106*, 6620-6625.
Sevcsik, E., Trexler, A.J., Dunn, J.M., and Rhoades, E. (2011). Allostery in a disordered protein: oxidative modifications to alpha-synuclein act distally to regulate membrane binding. Journal of the American Chemical Society *133*, 7152-7158.
Steffen, J.D., Tholey, R.M., Langelier, M.F., Planck, J.L., Schiewer, M.J., Lal, S., Bildzukewicz, N.A., Yeo, C.J., Knudsen, K.E., Brody, J.R.*, et al.* (2014). Targeting PARP-1 allosteric regulation offers therapeutic potential against cancer. Cancer research *74*, 31-37.
Stock, G., and Hamm, P. (2018). A non-equilibrium approach to allosteric communication. Philosophical transactions of the Royal Society of London. Series B, Biological sciences *373*.
Sun, J.H., O'Boyle, D.R., 2nd, Fridell, R.A., Langley, D.R., Wang, C., Roberts, S.B., Nower, P., Johnson, B.M., Moulin, F., Nophsker, M.J.*, et al.* (2015). Resensitizing daclatasvir-resistant hepatitis C variants by allosteric modulation of NS5A. Nature *527*, 245-248.
Swenson, D.W.H., Prince, J.H., Noe, F., Codera, J.D., and Bolhuis, P.G. (2018). OpenPathSampling: A Python framework for path sampling simulations. I. Basics. BioRxiv.
Taylor, N.D., Garruss, A.S., Moretti, R., Chan, S., Arbing, M.A., Cascio, D., Rogers, J.K., Isaacs, F.J., Kosuri, S., Baker, D.*, et al.* (2016). Engineering an allosteric transcription factor to respond to new ligands. Nat Methods *13*, 177-183.
Tee, W.V., Guarnera, E., and Berezovsky, I.N. (2018). Reversing allosteric communication: From detecting allosteric sites to inducing and tuning targeted allosteric response. PLoS computational biology *14*, e1006228.
Temiz, N.A., Meirovitch, E., and Bahar, I. (2004). Escherichia coli adenylate kinase dynamics: comparison of elastic network model modes with mode-coupling (15)N-NMR relaxation data. Proteins *57*, 468-480.
Tobi, D., and Bahar, I. (2005). Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. Proceedings of the National Academy of Sciences of the United States of America *102*, 18908-18913.
Toncrova, H., and McLeish, T.C. (2010). Substrate-modulated thermal fluctuations affect long-range allosteric signaling in protein homodimers: exemplified in CAP. Biophysical journal *98*, 2317-2326.
Tsai, C.J., and Nussinov, R. (2014). A unified view of "how allostery works". PLoS computational biology *10*, e1003394.
Vreede, J., Juraszek, J., and Bolhuis, P.G. (2010). Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. Proceedings of the National Academy of Sciences of the United States of America *107*, 2397-2402.
Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews. Molecular cell biology *16*, 18-29.
Xia, Y., DiPrimio, N., Keppel, T.R., Vo, B., Fraser, K., Battaile, K.P., Egan, C., Bystroff, C., Lovell, S., Weis, D.D.*, et al.* (2013). The designability of protein switches by chemical rescue of structure: mechanisms of inactivation and reactivation. Journal of the American Chemical Society *135*, 18840-18849.

Yan, L., Bouchaud, J.P., and Wyart, M. (2017a). Edge mode amplification in disordered elastic networks. Soft matter *13*, 5795-5801.
Yan, L., Ravasio, R., Brito, C., and Wyart, M. (2017b). Architecture and coevolution of allosteric materials. Proceedings of the National Academy of Sciences of the United States of America *114*, 2526-2531.
Yan, L., Ravasio, R., Brito, C., and Wyart, M. (2018). Principles for Optimal Cooperativity in Allosteric Materials. Biophysical journal *114*, 2787-2798.
Yifrach, O., and Horovitz, A. (1998). Mapping the transition state of the allosteric pathway of GroEL by protein engineering. J. Am. Chem. Soc *120*, 13262–13263.
Yoluk, O., Lindahl, E., and Andersson, M. (2015). Conformational gating dynamics in the GluCl anion-selective chloride channel. ACS chemical neuroscience *6*, 1459-1467.
Yuan, S., Filipek, S., and Vogel, H. (2016). A Gating Mechanism of the Serotonin 5-HT3 Receptor. Structure *24*, 816-825.
Zhu, L., Frenkel, D., and Bolhuis, P.G. (2011). Role of fluctuations in ligand binding cooperativity of membrane receptors. Physical review letters *106*, 168103.

Figure S1: Population shift between the inactive and active states of Ras-GDP/GTP.
(*Ruth Nussinov, Chung-Jung Tsa and Hyunbum Jang)*

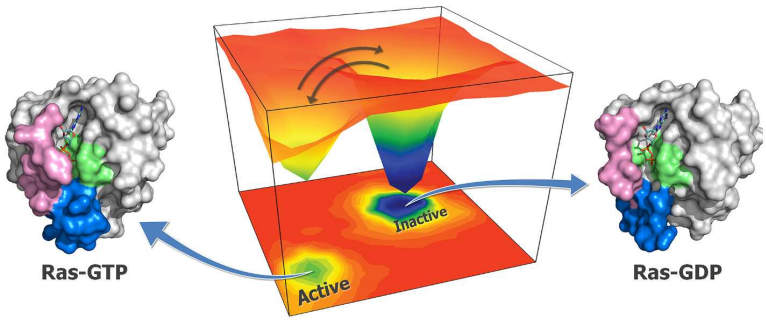


Illustration of the population shift phenomenon related to the authors' recent work on oncogenic mutants of the KRas4B isoform. Shown is the free energy landscape representing active Ras-GTP and inactive Ras-GDP states. Ras can be activated by guanine nucleotide exchange factors that exchange GDP with GTP. The active GTP-bound Ras can be inactivated through GTP hydrolysis by GTPase-activating protein (GAP).

References

-Tsai CJ, Nussinov R. A unified view of "how allostery works". PLoS Comput Biol 2014;10:e1003394
-Nussinov R, Tsai CJ. Unraveling structural mechanisms of allosteric drug action. Trends Pharmacol Sci 2014;35:256-264
-Ma B, Kumar S, Tsai CJ, et al. Transition-state ensemble in enzyme catalysis: possibility, reality, or necessity? J Theor Biol 2000;203:383-397

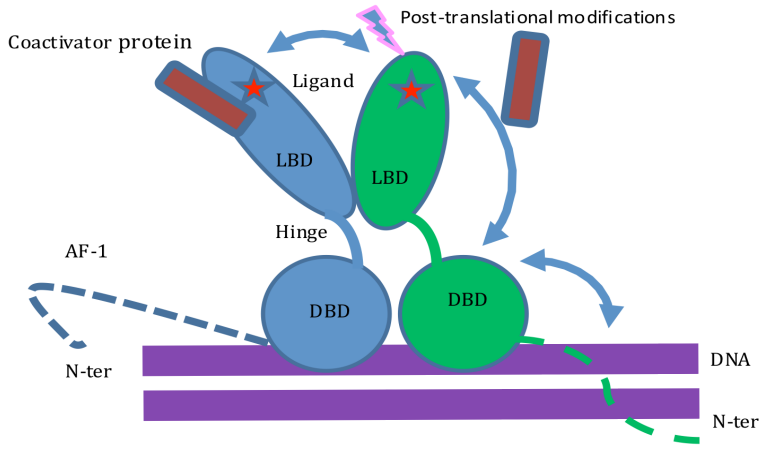


Figure S2: Allosteric regulation in nuclear receptors
(R.H. Stote, J. Eberhardt, Y. Chebaro, A. Dejaegere)

Schematic illustration of the domain structure of nuclear receptor proteins. This structure comprises the AF-1 N-terminal domain, DNA binding domain (DBD), hinge region, C-terminal ligand binding domain (LBD). Sources of allosteric signals include ligand binding and post-translational modifications, which can lead to allosteric signal transmission to the DBD affecting DNA binding. Ligand binding by one receptor can affect ligand binding by the second receptor and the DNA sequence can modulate co-activator binding.

Computational tools have been instrumental in understanding allosteric regulation of NRs [1],[2]. An early study on the PPARγ LBD identified networks of dynamically coupled amino acids that link the ligand binding pocket to the activation helix H12 and the heterodimer interface [3]. A more recently community network analysis has been applied to a larger PPAR architecture [4].

References

1. Grosdidier, S., et al., *Allosteric conversation in the androgen receptor ligand-binding domain surfaces.* Mol Endocrinol, 2012. **26**(7): p. 1078-90.
2. Burendahl, S. and L. Nilsson, *Computational studies of LXR molecular interactions reveal an allosteric communication pathway.* Proteins, 2012. **80**(1): p. 294-306.
3. Fidelak, J., et al., *Dynamic correlation networks in human peroxisome proliferator-activated receptor-gamma nuclear receptor protein.* Eur Biophys J, 2010. **39**(11): p. 1503-12.
4. Ricci, C.G., et al., *Allosteric Pathways in the PPARgamma-RXRalpha nuclear receptor complex.* Sci Rep, 2016. **6**: p. 19940.

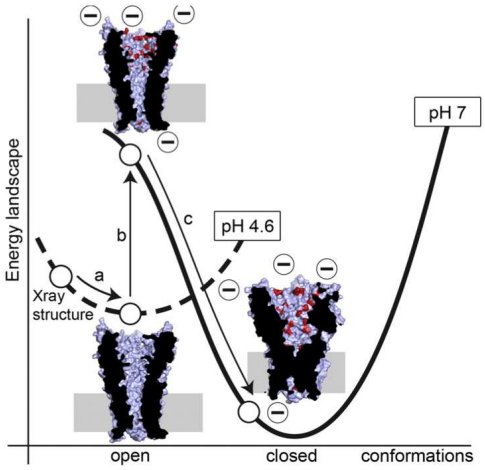


Figure S3: The Activation/relaxation MD approach to explore allosteric transitions of ligand-gated ion channels with atomic resolution: consequences for drug design
(Marco Cecchini and Jean-Pierre Changeux)

Schematic illustration of the activation/relaxation MD approach to allosteric transitions in ligand-gated ion channels. Starting from the X-ray structure of the bacterial receptor GLIC at pH 4.6 (open), a brief equilibration is carried out (a). Then, an instantaneous increase in pH is modeled (b), followed by the relaxation towards a closed conformation (c). The curves with broken and plain lines represent energy landscapes for pH 4.6 and 7.0, respectively. The protein's surface is represented in light blue with residues changing charge during the pH jump in red. Adapted with permission from Ref. [2].

The all-atom based molecular dynamics activation/relaxation approach described in the main text, relies on the hypothesis that allosteric proteins spontaneously populate a number of discrete conformational states in reversible equilibrium and that a conformational selection or *shift* of conformer population takes place upon ligand binding/unbinding [1]. In this view, modeling the addition (removal) of the agonist to (from) its binding site is expected to trigger a "vertical excitation" of the protein, whose structural relaxation by unbiased MD would capture the spontaneous transition to a distinct physiological state..

Lastly, knowledge of the transition path between pairs of physiologically relevant states at atomic resolution enables detailed analysis of the dynamics of the ligand-binding pockets. These include sites for orthosteric ligands and allosteric modulators, which are most often located at subunit or domains interfaces and typically open and close during the functional transitions between resting, active (and desensitized) states. Such analyses can be exploited for the rational design of positive or negative allosteric modulators [1], paving the way to new pharmacological strategies.

References

1. Cecchini, M. and J.P. Changeux, *The nicotinic acetylcholine receptor and its prokaryotic homologues: Structure, conformational transitions & allosteric modulation.* Neuropharmacology, 2015. **96**(Pt B): p. 137-49.

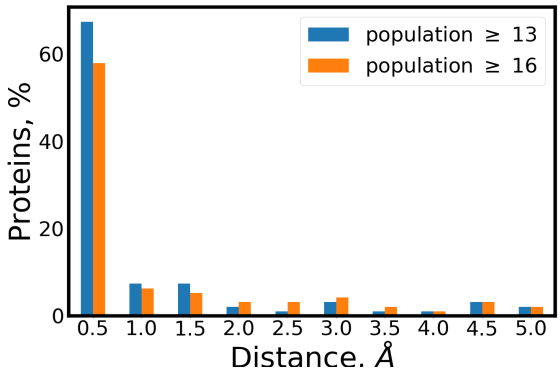


Figure S4: Structural Origins of Cryptic Binding Sites

(Dzmitry Padhorny and Dima Kozakov)

This figure shows the distribution of distances between the cryptic binding site and the closest FTsite probe cluster with above threshold population in the CryptoSite dataset [1]. Analysis is performed based on apo structures of the proteins. In the FTsite approach, clusters of size 13 and more are considered to be binding site predictors, which means that FTsite is able to correctly identify the vast majority of known cryptic binding sites. In a similar manner, clusters with more than 16 probes pinpoint "druggable" binding sites which could be targeted with sufficient affinity, thus showing that more than 80% of known cryptic binding sites are druggable.

References

1. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* **428**, 709–719 (2016).

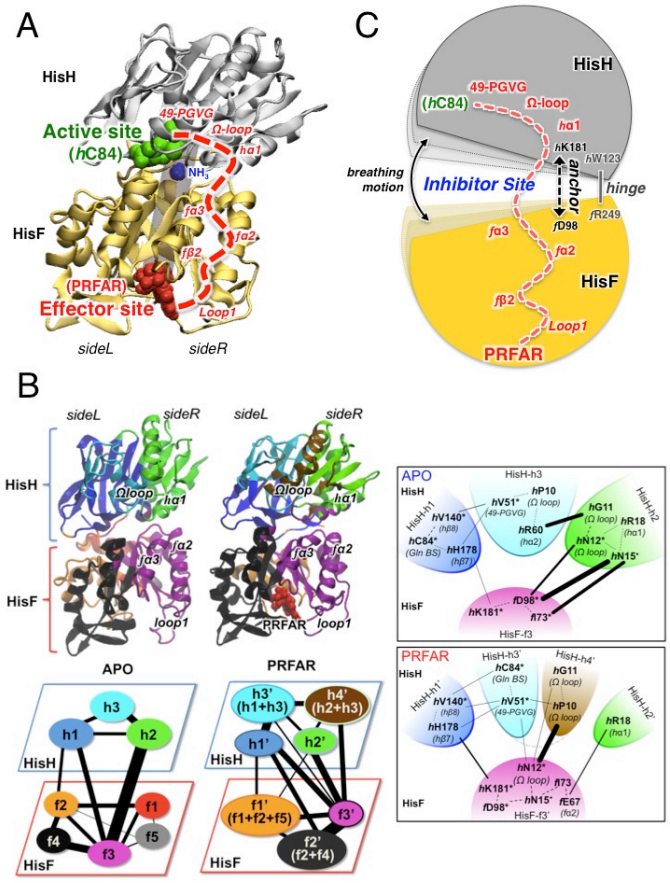


Figure S5: Community network analysis elucidating allostery regulation in the imidazole glycerol phosphate synthase (IGPS) protein
(Ivan Rivalta)

(A) Tertiary structures of *Thermotoga maritima* IGPS complex from *Thermotoga maritima*, with active site in the HisH protein (gray), allosteric site in the HisF protein (yellow) and ammonia (blue) channel (gray tube). IGPS catalyzes the hydrolysis of glutamine (Gln) at the *h* active site, upon binding of the effector PRFAR to HisF, >25 Å away. The allosteric pathways involve secondary structures (in red) on the IGPS right side (sideR). (B) Color-coded optimal community network of apo (left) and PRFAR-bound (right) complexes. The apo communities that contribute the most to the PRFAR communities are reported in brackets. Connectivity line-widths are proportional to the inter-communities total betweenness (ITB) with connections relevant to IGPS allostery depicted in the right panels. Fully and partially conserved residues are indicated with (*) and (•), respectively. Secondary structures are in brackets. (C) The allosteric communication involves structural changes induced by PRFAR binding, affecting the HisF/HisH breathing motion that has been targeted by small molecule inhibitors that bind at the interface. Community network analysis [2,3] was applied to derive optimal community networks of the apo and effector PRFAR-bound IGPS complexes (**B**), indicating that effector binding alters significantly the information flow between communities (thickness of the links between communities in **(B)**). The findings from these studies have been corroborated by measurements of NMR chemical shifts and were shown to be consistent with the expected inactive-to-active allosteric transition in IGPS. This studies have stimulated both single site mutagenesis experiments[4] and allosteric inhibitors design [5] (see **(C)**).

References

1. Manley G, Rivalta I, Loria JP (2013) Solution NMR and computational methods for understanding protein allostery. *Journal of Physical Chemistry B* 117 (11):3063-3073
2. Rivalta I, Sultan MM, Lee NS, Manley G, Loria JP, Batista VS (2012) Allosteric pathways in the imidazole glycerol phosphate synthase. *Proc Natl Acad Sci USA* 109:E1428-E1436
3. Sethi A, Eargle J, Black AA, Luthey-Schulten Z (2009) Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci USA* 106 (16):6620-6625
4. Lisi GP, East KW, Batista VS, Loria JP (2017) Altering the allosteric pathway in igps suppresses millisecond motions and catalytic activity. *Proc Natl Acad Sci USA* 114 (17):E3414-E3423
5. Rivalta I, Lisi GP, Snoeberger N-S, Manley G, Loria JP, Batista VS (2016) Allosteric communication disrupted by a small molecule binding to the imidazole glycerol phosphate synthase protein–protein interface. *Biochemistry* 55 (47):6484-6494

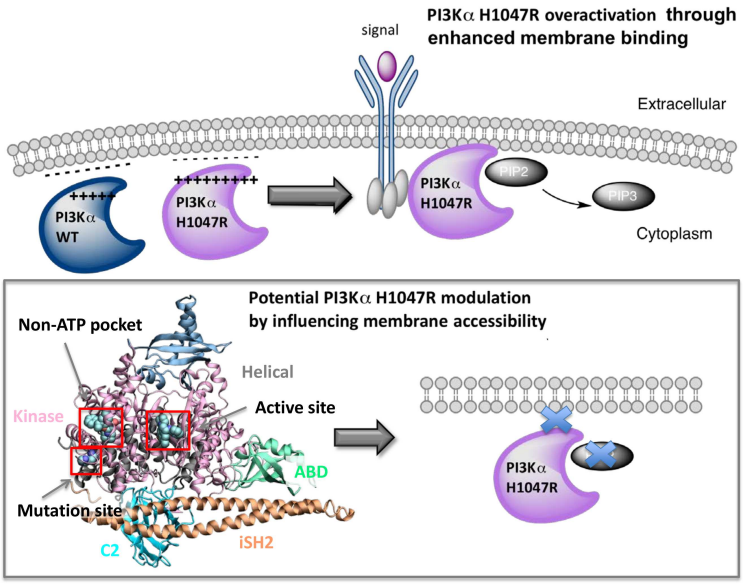


Figure S6: Allosteric modulation in drug discovery: The PI3Kα paradigm

(Paraskevi Gkeka)

Schematic illustration of the proposed allosteric mechanism phosphoinositide 3-kinase alpha (PI3Kα)[1,2].

PI3Ks are lipid kinases that play key roles in many fundamental biological processes, including cell growth, proliferation, differentiation, motility, survival, and metabolism. Among the different PI3K isoforms, PI3Kα is the most important as it plays a pivotal role in cell proliferation in response to growth factor-tyrosine kinase pathway activation [3]. PI3Kα is a heterodimer that contains a p110α catalytic subunit and p85α regulatory subunit [1], and is a target of particular pharmacological interest for anti-cancer drug development [3,4]. Currently, two PI3Kα inhibitors, Alpelisib and MLN1117, are in Phase III trial for patients with HR+/HER2–advanced breast cancer and Phase II trial in patients with advanced solid tumors, respectively (https://clinicaltrials.gov, November 2017). Despite the advances in targeting the PI3Kα isoform, a major challenge remains because of the highly conserved ATP binding sites of lipid kinases. Based on molecular modeling techniques and surface plasmon resonance experiments, a multifactorial model of the over-activation mechanism of the most common *PIK3CA* mutant (H1047R), was proposed [1]. Using this model a new binding pocket distinct from the active site, and close to residue 1047,was identified. This pocket was further explored for potential allosteric modulation of the H1047R PI3Kα mutant [2]. Positional covariance, protein fluctuation analyses showed that the main areas involved in protein conformational changes upon ligand binding to the non-ATP pocket are the membrane binding domains. PI3Kα mutant activity could therefore, be modulated not in terms of the active site activity, but by altering the dynamics of protein-cell membrane interaction and subsequent substrate retrieval. Thus, the non-ATP pocket could potentially be used for the discovery of a selective inhibitor of protein−membrane interactions tailored for the H1047R mutant protein. Such an inhibitor could act by constraining PI3Kα membrane-binding motifs, influencing membrane accessibility and subsequent substrate availability, representing a promising alternative or complementary strategy for allosteric PI3Kα modulation.

References

[1] Gkeka, P; Evangelidis, T; Pavlaki, M; Lazani, V; Christoforidis, S; Agianian, B; Cournia, Z: Investigating the Structure and Dynamics of the *PIK3CA* Wild-Type and H1047R Oncogenic Mutant. *PLoS Comput Biol* **2014**, 10(10): e1003895.

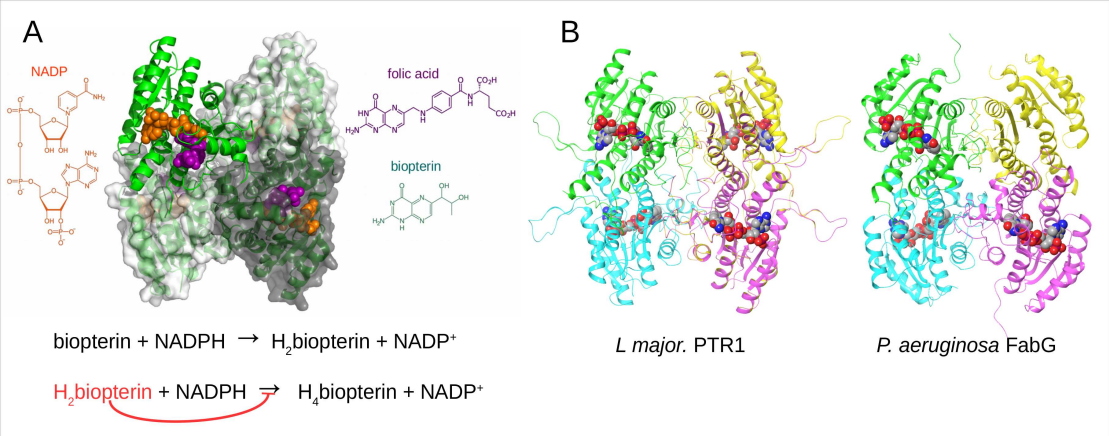
[2] Gkeka P; Papafotika, A; Christoforidis, S; Cournia, Z: Exploring a Non-ATP Pocket for Potential Allosteric Modulation of PI3Kα. *J. Phys. Chem. B* **2015,** 119 (3), 1002-1016.

[3] Samuels, Y. *et al*. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **2004**, 304:554.

[4] Fruman, D A.; Chiu, H; Hopkins, B D; Bagrodia, S; Cantley, L C; Abraham, R T: The PI3K Pathway in Human Disease Cell **2017**, 170(4), 605–635.

Figure S7: Allosteric modulation of pteridine reductase 1, a member of the short-chain dehydrogenases/reductases family

(Joanna Panecka-Hofman and Rebecca C. Wade)



Displayed are experimental and modeled structures of Pteridine reductase 1 (PTR1, EC 1.5.1.33), a folate pathway enzyme unique to trypanosomatid parasites. (**A**) PTR1 homotetramer structure (*T. brucei*, pdb: 3bmc), and the chemical structures of the main substrates and the NADP cofactor, and the reaction catalyzed by PTR1 with biopterin as an exemplary substrate (substrate inhibition shown in red). (**B**) Structures of PTR1 (a model based on pdb: 1e92) and FabG (pdb: 4ag3). The proteins are shown in cartoon and molecular surface representation and NADP as van der Waals spheres (only heavy atoms are shown). PTR1 catalyzes the reduction of folate and biopterin with NADPH cofactor [1]. It belongs to a large family of short-chain dehydrogenases/reductases (SDRs), which feature diverse sequences and very diverse functions [3], but have a highly conserved Rossmann fold [3]. PTR1 is a potential target for anti-parasitic drugs, but existing inhibitors require optimization [2]. Targeting allosteric sites of PTR1 is therefore an attractive alternative.

PTR1 is inhibited by semi-products of its catalytic reaction (substrate inhibition) (**A**)) at substrate concentration of ~10-mM levels, and may involve an allosteric mechanism [1]. Cooperativity between the binding sites is also postulated for other SDR enzymes, including homotetrameric bacterial 3-ketoacyl-[acyl-carrier-protein] reductase (FabG, **B**)[4], and allosteric inhibitors binding at the inter-subunit interfaces were reported for FabG [5]. The FabG tertiary and quaternary structure is similar to that of PTR1, despite low sequence identity (~29% for *L. major* PTR1 and *P. aeruginosa* FabG (**B)**). The possibility of allosteric regulation in PTR1 was investigated using normal mode analysis (NMA), indicating significant concerted movements of the substrate loops flanking the active site. In addition, Rotamerically Induced Perturbation simulations (RIP [6]), revealed flexibility hot-spots in the PTR1 homotetramer structure, and suggested dynamical allosteric couplings between distant protein residues.

References

[1] Nare B, Hardy LW, Beverley SM. J Biol Chem. 1997; 272:13883-13891. doi: 10.1074/jbc.272.21.13883

[2] Panecka-Hofman J, Pöhner I, Spyrakis F, Zeppelin T, Di Pisa F, Dello Iacono L, Bonucci A, Quotadamo A, Venturelli A, Mangani S, Costi MP, Wade RC. Biochim Biophys Acta. 2017; 1861(12):3215-3230. doi: 10.1016/j.bbagen.2017.09.012.

[3] Kavanagh KL, Jörnvall H, Persson B, Oppermann U. Cell Mol Life Sci. 2008 Dec; 65(24):3895-906. doi: 10.1007/s00018-008-8588-y.

[4] Price AC, Zhang YM, Rock CO, White SW. Biochemistry. 2001 Oct; 40(43):12772-81.

[5] Cukier CD, Hope AG, Elamin AA, Moynie L, Schnell R, Schach S, Kneuper H, Singh M, Naismith JH, Lindqvist Y, Gray DW, Schneider G. ACS Chem Biol. 2013 Nov;8(11):2518-27. doi: 10.1021/cb4005063.

[6] Ho BK, Agard DA. PLoS Comput Biol. 2009 Apr; 5(4):e1000343. doi:10.1371/journal.pcbi.1000343.

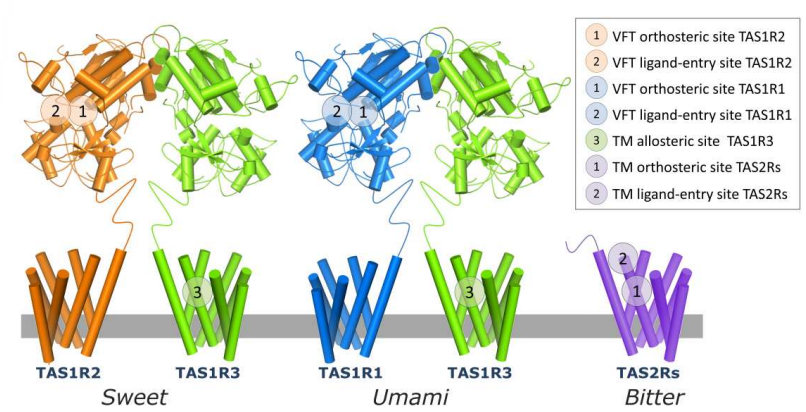


Figure S8: Allosteric modulation of taste GPCRs
(Antonella Di Pizio and Masha Y. Niv)

The Figure illustrates the ligand-binding sites of sweet (TAS1R2/TAS1R3, in orange and green), umami (TAS1R1/TAS1R3, in blue and green) and bitter (TAS2Rs, in violet) taste receptors, members of the G-protein coupled receptors (GPCRs) family.

The odorant or olfactory receptors (ORs) represent the largest family of GPCRs, with over 1000 members in mice and ~400 in humans. Sweet (elicited by sugars) and umami (elicited by glutamate) taste modalities are mediated via TAS1Rs, and bitter (elicited by caffeine, quinine, strychnine and hundreds of other compounds), is mediated by the TAS2Rs family that counts 25 subtypes in humans. Only the crystal structures of the extracellular VFT regions of the fish Tas1r2/Tas1r3 heterodimer are currently available [1], but iterative combinations of simulation and experiment have been successfully used for delineating binding modes of tastants, and for predicting additional ligands [2]. TAS1Rs (Class C GPCRs) contain an extracellular VFT as an N-terminal domain, linked to the TM domain via a short cysteine-rich (CR) domain. VFT, the orthosteric ligand-binding domain, consists of two lobes that can assume an open (inactive) or closed (active) conformation. Umami taste is mediated by a heterodimer composed of TAS1R1 and TAS1R3, and sweet is sensed by a heterodimer of TAS1R2 and TAS1R3.

The Sweet receptor, TAS1R2/TAS1R3 heterodimer is activated by diverse compounds, ranging from low molecular weight sweeteners to sweet proteins. The umami receptor, TAS1R1/TAS1R3 heterodimer, specifically responds to glutamate. Integrated computational approaches and mutagenesis studies identified the binding site for sugars in the VFT domain of TAS1R2, and of glutamate in the VFT domain of TAS1R1, as reviewed in [2]. An allosteric binding site for small molecules has also been identified in the TM domain of TAS1R3. Cyclamate and lactisole bind to an allosteric site in the TM domain, and act as allosteric enhancer and negative modulators, respectively, for both sweet and umami receptors [2]. In addition, multiple binding sites for different ligands have been identified on the sweet taste receptor. For example, small agonist molecules, such as sucralose, saccharin, aspartame and neotame, bind to the orthosteric binding site in the TAS1R2 VFT domain, but positive allosteric modulators, sweet taste enhancers, were found to bind at the entrance of the VFT domain of the TAS1R2 (ligand-entry site).

Allosteric regulation in bitter receptors is less well understood. TAS2Rs are classified as Class A GPCRs, although the sequence similarity is very low (13-29% for the TM domains). Despite low sequence identity the orthosteric binding pocket of bitter taste receptors coincides with the canonical binding site of Class A GPCRs. In addition, an extracellular or vestibular site has been described to be transiently involved in the binding process [3].

References

1. Nuemket N, Yasui N, Kusakabe Y, Nomura Y, Atsumi N, et al. (2017) Structural basis for perception of diverse chemical substances by T1r taste receptors. Nat Commun 8: 15530.
2. Di Pizio A, Niv MY (2014) Computational Studies of Smell and Taste Receptors. Israel Journal of Chemistry 54: 1205-1218.
3. Di Pizio A, Levit A, Slutzki M, Behrens M, Karaman R, et al. (2016) Comparing Class A GPCRs to bitter taste receptors: Structural motifs, ligand interactions and agonist-to-antagonist ratios. Methods Cell Biol 132: 401-427.