NMR tools to detect protein allostery

Journal Article

Author(s):
Gampp, Olivia; Kadavath, Harindranath; Riek, Roland

Publication date:
2024-06

Permanent link:
https://doi.org/10.3929/ethz-b-000663720

Rights / license:
Creative Commons Attribution 4.0 International

Originally published in:
Current Opinion in Structural Biology 86, https://doi.org/10.1016/j.sbi.2024.102792

Funding acknowledgement:
219645 - On the Structural Landscape of Proteins (SNF)
NMR tools to detect protein allostery
Olivia Gampp¹, Harindranath Kadavath¹,² and Roland Riek¹

Abstract
Allostery is a fundamental mechanism of cellular homeostasis by intra-protein communication between distinct functional sites. It is an internal process of proteins to steer interactions not only with each other but also with other biomolecules such as lipids, lipids, and nucleic acids. In addition, allosteric regulation is particularly important in enzymatic activities. A major challenge in structural and molecular biology today is unraveling allosteric sites in proteins, to elucidate the detailed mechanism of allostery and the development of allosteric drugs. Here we summarize the recently developed tools and approaches which enable the elucidation of regulatory hotspots and correlated motion in biomolecules, focusing primarily on solution-state nuclear magnetic resonance spectroscopy (NMR). These tools open an avenue towards a rational understanding of the mechanism of allostery and provide essential information for the design of allosteric drugs.

Addresses
¹ Laboratory of Physical Chemistry, ETH Zurich, Switzerland
² St. Jude Children’s Research Hospital, 262 Danny Thomas Place, 38105 Memphis, Tennessee, USA

Corresponding author: Riek, Roland (roland.riek@phys.chem.ethz.ch) (Kadavath H.)

Introduction
Protein allostery has been a subject of research since the 1960s when the term was first coined to describe the interaction of the multimeric protein hemoglobin with its allosteric partner, oxygen [1]. Over the years, the concept of allosteric regulation has evolved significantly. It has transitioned from the initial models proposed by Monod (MWC model) and Koshland (KNF model) to the contemporary understanding that all proteins exhibit some form of allosteric behavior [2–6]. This can encompass alterations in protein structure, such as the conformation of a multimeric protein, or changes in the dynamics within a single protein domain including correlated and concerted motion. Therefore, current models such as the ensemble allosteric model (EAM) has expanded on the original models to describe these findings [7].

Understanding the mechanical intricacies of allosteric phenomena remains a challenge, and pinpointing the specific residues involved in allosteric networks has proven to be difficult [8]. This challenge may arise from the fact that, under physiological conditions, proteins undergo numerous conformational changes, resulting in a wide spectrum of motions ranging from $10^{-10}$ to $10^{-7}$ m, with durations spanning from $10^{-12}$ s to $10^{-5}$ s or even longer. While it is believed that virtually every protein exhibits some form of allostery, only a few thousands of proteins have been identified as such [9].

This short review primarily focuses on novel tools and software designed for the identification and quantification of correlated motions within proteins at atomic resolution. Correlated motion describes the concerted exchange of an extensive network of atoms in a protein between structural states. Since in correlated motion distinct sites of a protein move together, they are linked with each other, and it is thus obvious that it may serve as a basis for protein allostery. We begin by offering an overview of advances in nuclear magnetic resonance (NMR) methodology that enable the atomic-resolution determination of multiple states of three-dimensional protein structures. The strength of atomic resolution relies on the quantitative nature of the coordinates allowing the elucidation of the origin of protein allostery. We also provide a concise summary of lower-resolution NMR information that can be harnessed to investigate allosteric phenomena. Finally, we compile an array of available tools for revealing correlated protein motions, comparing their limitations and suitable use cases. Thus, the review presented here is considered more focused compared to the recent comprehensive review by Ramelot et al. recently published in the same journal, which also encompasses a wide array of predictive tools [10].

Allostery at atomic resolution: multi-state protein structure determination by NMR
When allostery was initially described in hemoglobin and myoglobin, X-ray crystallography served as the
primary method for determining their structures, yielding resolutions of 5.5 and 2 Å, respectively [11–14]. Since then, the experimental determination of three-dimensional structures has remained a fundamental practice for gaining insights into protein allostery. With the leaps made in technology to produce and detect X-rays, e.g. synchrotrons and photon counting detectors, X-ray crystallography remains not only a suitable method to probe the structure but also to study the dynamics of a biomolecule [15,16]. In addition to X-ray crystallography new techniques and methods have been developed to obtain not only high-resolution three-dimensional structures, but also sets of structures of a single protein in distinct environments, such as different crystals, buffers, ligands, and even in the presence of distinct electric fields, as well as the elucidation of multiple states within a single environment which have furthered our understanding of allostery [17–31].

Since solution-state NMR is measured under physiological conditions in aqueous solution, the development of NMR spectroscopy has played a pivotal role in our current understanding of protein allostery. In addition to atomic resolution information of the protein structure, NMR also provides insights into dynamics and thermodynamic properties based on probes within the protein, such as NMR active 1H, 13C and 15N, that are ensemble- and time-sensitive. Consequently, it stands out as one of the most suitable experimental methods for the detection and analysis of allosteric phenomena [26–29,32–39]. Already the simple chemical shift perturbation of a ligand titration typically recorded in a [15N,1H]-HSQC experiment may reveal the presence of protein allostery (see below). There are however much more sophisticated NMR methods available including residual dipolar couplings (RDCs) [40], chemical shifts [37,41], relaxation dispersion [42,43], cross-correlated relaxation (CCR) [44], paramagnetic relaxation enhancement (PRE) [45,46], methyl order parameters [47], chemical exchange saturation transfer (CEST) [48–50] and exact Nuclear Overhauser Enhancement or Effect (eNOE) in combination with molecular dynamics simulation, structure prediction software, or ensemble-based structure calculations [51–54]. These probes show different sensitivity and selectivity. While some of the methods are sensitive to low populated states (such as relaxation dispersion) or to rates (such as relaxation measurements), the advantage of the eNOE approach of interest is resolving correlated structural states at atomic resolution by multi-state protein structure determination.

**eNOE method to dissect protein allostery**

The standard procedure to determine a protein's structure using NMR established in the 1980s is based on the so-called nuclear Overhauser enhancement (NOE) phenomenon [55], as the NOE rate is inversely proportional to the ensemble average of the sixth power of the distance between two interacting spins yielding thousands of distances between 1H. With the introduction of the exact NOE (eNOE) method, distances between 1H with an accuracy of 0.1 Å can be obtained [51,56], making it together with RDCs one of the most high-resolution structure elucidation technique of biomolecules today. The high precision achieved through eNOEs enables the separation of conformations that are otherwise averaged into a single state by the nature of the calculation. This leads to the possibility of employing ensemble-based structure calculations on samples in thermal equilibrium, as opposed to the standard procedure for calculating single-state models. The multi-state structure calculations can be done using the eNORA tool [51,57], which is implemented in the program CYANA [58].

To date, the method was able to demonstrate that several proteins and RNA in solution exist in multi-states that interchange in the micro-second time range [32]. Usually, two distinct states are identified, while higher state structure calculations are also possible if the data are of high quality [59]. The separation of the states at atomic resolution has provided the allosteric coupling mechanisms at atomic resolution related to biomolecule’s function [27,60–62]. In the WW-domain of PIN1, a positive allosteric ligand moves the population between two states known as conformational selection allostery, while a negative allosteric ligand (allostery-suppressing) interferes with the correlated dynamics yielding anti-correlated dynamics. The WW-domain of PIN1 is a prime example of dynamic allostery, in which conformational changes are very subtle during the allosteric regulation that traditional structure determination methods would likely not detect [63].

Finally, in greater detail, the three-level allosteric network of apo and holo forms of the PDZ2 domain of human tyrosine phosphatase 1E (chPTP1E) and corresponding two-state structures are discussed since they can be corroborated with the findings from three other separate methods. Those being firstly evolutionary data, which took the entire family of PDZ domains instead of just hPTP1E to infer conserved allosteric interactions such as ligand binding; secondly, MD simulations uncovering dynamics taking place on a much shorter timescale, due to the length of the typical MD trajectory being roughly 2 ns long; and lastly, NMR relaxation data, which has limitation that it assumes regions with similar rates belong to the same network. In Figure 1 the allosteric networks obtained from the multi-state calculations (a-c) are compared to the three different methods (d-f). The networks obtained from the eNOE-based apo PDZ2 structures resemble the ones deduced from the methyl relaxation data (f) and MD data (e). On the other hand, the evolutionary data-derived network (d) resembles the eNOE-deduced allosteric network of
binding. The multi-state structure of the PDZ2 domain displays the method’s ability to elucidate different structural aspects that can be interpreted as protein allostery at atomic resolution. However, the dynamic nature of the protein can be further investigated with kinetic or relaxation measurements [64].

Evaluation of correlated motion from structural ensembles

Once multi-state structures from ensembles have been derived, whether this was done with experimental data as described above or in silico ensembles collected from umbrella or other elevated sampling techniques within molecular dynamics (MD) simulations, it may be surprising to the reader, that it is not straightforward to elucidate structural correlation and determine if it is of allosteric nature, without subjective alignment of the ensembles. From the MD field, there are two general methods used to extract correlations. The first being normal mode analysis, and the second, principal component analysis (PCA) on the atomic covariance matrix. Normal mode analysis assumes that the functionally relevant motions in a biomolecule have large amplitudes and that these correspond to the vibrational normal modes with the lowest frequencies. Whereas PCA reduces the dimensionality of the computed covariance matrix after superposition. There have been a plethora of tools, R and python packages that have been developed which use these two methods to determine allosteric sites in proteins. The most common ones are correlationplus [65], MDAnalysis [66], NAPS [67], SINAPs [68], WebPSN [69], Bio3D [70], RING-PyMOL [71], PyInteraph2 [72], Ohm [73] and gRINN [74]. They all rely on various formats of MD trajectories. Recently gRINN (get Residue Interaction eNergies and Networks) enabled the discovery of the allosteric network in an adenosine A1 receptor [75]. The tool WISP (Weighted Implementation of Suboptimal Paths) takes MD trajectories in the PDB format and allows the user to determine the optimal path as well as suboptimal paths [76].

Recently the software package/server PDBcor was introduced which analyses an ensemble of conformations in PDB format such as conformer bundles typically representing NMR or MD trajectories/ensembles [77]. PDBcor uses either the torsion angles or distance statistics to extract correlations without superimposing the
conformers as needed with principal component analysis. With this objective the distance between any 2 residues or the torsion angles in each protein conformer are clustered using a Gaussian mixture model (GMM), then using mutual information, a method from information theory, a correlation is extracted for every residue pair. This yields a quantitative localization and detection of correlated motion as demonstrated in Figure 1.

**NMR methods to detect allostery at residue resolution**

As mentioned above, a powerful straightforward low-resolution technique to study protein allostery is ligand-induced chemical shift perturbation analysis, which enable the investigation of allosteric interactions that not only alter the protein's structure but also impact its dynamics at residue resolution (i.e. typically one $^{15}\text{N}-^{1}\text{H}$ moiety per amino acid residue is monitored). These displacements are highly sensitive to even minor structural or dynamical changes, and the peak positions provide valuable information regarding the ensemble’s conformer populations. A recent example of this kind is the very well-studied and known *Escherichia coli* Lac repressor [41]. In the absence of ligands, the dimer exists in a dynamic equilibrium between DNA-bound and inducer-bound conformations. In the ternary complex between inducer, repressor, and operator DNA, induced-allosteric changes disrupt the interdomain contacts between the inducer binding and DNA binding domains releasing the Lac repressor from the operator following the MWC model.

The analysis of ligand-induced chemical shift perturbations was put to another level of sophistication. The CHESCA tool (CHEmical Shift Covariance Analysis) was developed to establish allosteric networks based on the chemical shift displacements observed between the spectra of the apo and effector-bound states [35,78]. It operates under the assumption that all residues belonging to the same allosteric network experience perturbations upon effector/ligand binding, resulting in variations in the chemical shifts of these residues. To identify and classify the residues partaking in the network, CHESCA combines agglomerative clustering (AC), to group coupled residues, and singular value decomposition (SVD) to evaluate if a residue is binding the effector/ligand or is allosteric. The model was validated on several known allosteric proteins [79] and applied to multiple other systems since then [80]. CHECSA has even been used on solid-state NMR data of the ion channel KcsA reconstituted in (DOPE/DOPS) liposomes. Since this potassium channel binds protons in addition to potassium and these ligand binding events are allosterically coupled, the CAP method (chemical shift detection of allostery participants) was developed to identify residues that are potentially mediating the allosteric coupling [81]. Furthermore, the allosteric site found in the protein CheY using MD trajectories analyzed by Ohm could be substantiated with NMR experimental data evaluated using the CHESCA tool [82].

Carr-Purcell-Meiboom-Gill (CPMG) and chemical exchange by saturation transfer (CEST) NMR experiments are particularly useful for detecting allostery experimentally, since they provide information on chemical exchange, conformational dynamics taking place on the intermediate timescale, and low-populated excited states that are not visible using other methods. The experiments can be done on proteins with several hundreds of kDa size, thanks to the Methyl-TROSY-based $^{1}\text{H}$ CPMG [83]. For example, residues involved in the allosteric network of the human thymidylate synthase (hTS), a target for cancer therapies, could be discovered due to experimental data from methyl-based CPMG and CEST NMR. From the data, the authors discovered a novel form of dynamic allostery that was not detectable from crystal structures [50]. The combination of CPMG and CEST experiments was further able to show that the dimeric allosteric protein yeast chorismate mutase (CM) switches between MWC and KNF depending on the presence or absence of tryptophan, and its observed allosteric mechanisms can only be fully described by the ENM model [84].

Recently the design of a protein was improved by using $^{15}\text{N}$- and $^{1}\text{H}$-CEST experiments to obtain data on the apo form of C34 protein, an $\alpha/\beta$ protein that binds to specific regions in amyloidogenic proteins such as tau and A\beta42. The apo form of the protein is highly dynamic and does not reflect the ordered structure predicted by deep-learning protein structure prediction algorithms with high confidence. The broadened peaks due to dynamics could be accessed with the CEST experiments allowing the two closed conformations between which the apo-form switches to be studied and this data was exploited when designing the new variant of C34 of which the affinity for the octapeptide binding partner is now allosterically modulated by the binding of a second ligand [85].

**Conclusion**

With the recent advances in multi-state structure determination and dynamics studies using solution-state NMR, the origin of protein allostery at atomic resolution is no longer enigmatic. As evidenced by the selected systems studied in thermal equilibrium in solution, the mechanistic origin of allostery appears to be multifaceted. This demands establishing new experimental methods to detect correlated motion for many protein systems, primarily because, the structure-dynamics-function paradigm of proteins is far from complete and further because correlated motion and multi-state ensembles cannot be predicted, meaning that in silico
structure elucidation by AlphaFold is uninformative [53, 85, 86]. Furthermore, it is expected that Darwinian evolution optimized the systems to great detail well below the $k_BT$ range above which molecular dynamics simulation may have some predictive power. In addition, it is stated that the experimental determination of multi-states of a protein at atomic resolution is believed to be one of the most powerful tools in quantitative biology far above some drawings derived from low-resolution data such as chemical shift perturbation or dynamic rate determination, because of the quantitative nature of the data stored and represented as coordinates.

Declaration of competing interest
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability
No data was used for the research described in the article.

Acknowledgments
This work was funded in part by the Swiss National Science Foundation (SNSF) SNF 205320_219645/1.

References

Papers of particular interest, published within the period of review, have been highlighted as:
*  of special interest
**  of outstanding interest


The standard reference that introduced the conformational selection model to explain protein allostery.


With the use of an electric field the correlated structural network of a protein could be investigated by x-ray crystallography.


39. Comprehensive, well-illustrated review on NMR methods to study dynamic allosterie.


78. Straight forward elucidation of allosteric networks using chemical shift perturbation studies with ligands having similar activities.


85. The Methyl-TROSY-Based 1H Relaxation allows the elucidation of conformational dynamics by solution state NMR on large proteins.

85. Rennella E, Sahtoe DD, Baker D, Kay LE: **Exploiting conformational dynamics to modulate the function of designed proteins.** *Proc Natl Acad Sci U S A* 2023, **120**, e2303149120, https://doi.org/10.1073/PNAS.2303149120/SUPPL_FILE/PNAS.2303149120.SAPP.PDF.