A Minimal Approach to the Dynamic Regulation of Biomolecular Conformation

Dissertation
for the award of the degree
"Doctor rerum naturalium" (Dr. rer. nat)

of the
Georg-August-Universität Göttingen
within the doctoral program
Physics of Biological and Complex Systems
of the
Georg-August University School of Science

submitted by

Maximilian Alexander Vossel

from Willich

Göttingen, 2022

Thesis Committee

First referee and supervisor

Dr. Aljaž Godec
Mathematical bioPhysics Group
Max-Planck-Institut für Multidisziplinäre Naturwissenschaften, Göttingen

Second referee

Prof. Dr. Peter Sollich
Non-equilibrium Statistical Physics, Institut für Theoretische Physik
Georg-August-Universität Göttingen

Prof. Dr. Matthias Krüger
Non-equilibrium Statistical Physics, Institut für Theoretische Physik
Georg-August-Universität Göttingen

Examination Board

Dr. Aljaž Godec

Prof. Dr. Peter Sollich

Prof. Dr. Matthias Krüger

Prof. Dr. Stefan Klumpp
Theoretische Biophysik, Institut für Nichtlineare Dynamik
Georg-August-Universität Göttingen

Prof. Dr. Jörg Enderlein
Biophysik / Komplexe Systeme, III. Physikalisches Institut
Georg-August-Universität Göttingen

Dr. Andreas Neef
Labor Neurophysik
Göttingen Campus Institut für Dynamik biologischer Netzwerke

Date of oral examination: July 15, 2022

Dedication

To Carolin

Acknowledgements

First of all, I would like to express my gratitude to my supervisor Aljaž Godec, who entrusted me with a wonderfully fascinating and complex project and always believed that we would find a new answer to this old question. You are impressive and inspiring in so many aspects. I will never forget your enthusiasm, patience, and encouragement throughout the good and the challenging times. I wish you all the brightest for your scientific and personal future and hope your door will always stay open.

I am grateful for many valuable discussions with Peter Sollich and Matthias Krüger. Your sharp minds and quick physical perception helped me sift through the weeds of details and filter out the essential questions.

I want to thank all the *Mathematical bioPhysics* group members, particularly David Hartich, Kristian Blom, Alessio Lapolla, Gerrit Wellecke, and Cai Dieball. It was always fun and exciting to work with you, just as science should be. All of you always had an open ear for me. I am especially grateful for many valuable discussions with Cai Dieball; your critical reasoning was helpful and inspiring. Kris, thank you for the great atmosphere in the office, it was a great time.

Bert de Groot, I am very grateful for your suggestion to simply attempt to use our findings to predict the allosteric sites; it has proven valuable. I want to thank Lars Bock and Wojciech Kopec for your support in biology-related questions. I also want to thank Helmut Grubmüller for providing the infrastructure and attracting me to biophysics in Göttingen with an inspiring talk at the physics colloquium of the University Duisburg-Essen (now a long time ago).

I will remember the whole department of Theoretical and Computational Biophysics for an always helpful and honest atmosphere; I especially enjoyed our lunch-cooking-group. I would also like to acknowledge Frauke Bergmann, Antje Erdmann, Eveline Heinemann, Stefanie Teichmann and Petra Kellers for all their support in administrative issues. I want

to further thank Ansgar Esztermann, Martin Fechner, and Carsten Kutzner for the great support regarding hard- and software, not limited to the office equipment. Carsten, the coffee tree already produced the first beans.

Then there are of course Daniel Härtter, Elsa Steinfath, Sarah Henze und Jost Kollmeier. Although we share a passion for science, that is not what I primarily want to thank you for. You made my time in Göttingen great fun in so many aspects. It would have been dull without you. I am indebted to Max Freudenberg and Lukas Engelke. You have been available for my numerous technical questions at any time of the day (or night). Most importantly, you have asked me the right questions.

I would like to thank my family, especially my parents, for their trust, love, and unconditional support they gave me throughout the time of the thesis and also in the years of studying before. You made this possible.

Ada, thank you for just being you. Whatever might have gone wrong at work, when I got home, and you came running towards me, everything else was forgotten.

Last but not least, I am incredibly thankful to you, Carolin. What you have done in the past weeks, months, actually years is incredible. I could never have done it without you. Thank you.

Abstract

Allostery is the key mechanism enabling a “remote” regulation of the activity of proteins and other biological macromolecules. It forms the basis for controlling various inter- and intracellular processes. Allosteric proteins accommodate changes at their active site as a result of ligand binding at a second, often spatially distant binding site—the allosteric site.

Despite the ubiquity and fundamental role of allostery in proteins, the general physical principle underlying the efficient and precise long-range transmission of mechanical signals from the allosteric to the active site remains elusive. In addition to contributing to the unraveling of fundamental processes essential for sustaining life, the elucidation of a general mechanism holds a significant practical potential: It allows the prediction of allosteric sites in proteins based on single structures and hence addresses the current bottleneck for allosteric drug design.

Here, building on the previous success of spectral methods in explaining the dynamics of proteins, we hypothesize a simple physics-based principle that unifies allosteric behavior in trained artificial and protein-derived elastic networks. The mechanism rests on the concept of a collective lever that couples the interaction of stiff and soft modes in a sophisticated way. Input displacements at allosteric source pockets efficiently load collective stiff springs while the response occurs via soft modes, which convert the stored energy into large, nonlocal but specific displacements at the target pocket.

To test this hypothesis, we develop a fast and accurate algorithm for determining the full (nonlinear) response of elastic networks to perturbations caused by displacements at the source pockets and employ it to evolutionarily train networks to display allosteric responses. We observe nonlinear and non-reciprocal behavior during the responses and discuss the origin and the implications this may have for allostery.

Using spectral and perturbative approaches, we find convincing evidence confirming the predicted properties that characterize the uniqueness of the real source compared to other possible binding pockets in protein-derived and artificial, trained networks.

Finally, we demonstrate the applicability of the concept in drug design by using it to predict allosteric source pockets in proteins—with remarkable success, considering that only the initial structure and a single-parameter model is employed.

These findings shift the paradigm of allosteric signal propagation in networks and proteins from a purely soft-mode based interpretation towards a two-step thinking, which

differentiates strongly between the allosteric in- and output. We conclude that allostery is only fully comprehensible if viewed as what it actually is—a non-equilibrium effect that requires the perturbation to be accounted for explicitly in the spectral interpretation.

Keywords *allostery, proteins, conformational motion, elastic network, collective modes, mode-coupling, nonlinearity, non-reciprocity, evolutionary optimization, network design, drug design*

Zusammenfassung

Allosterie ist der Mechanismus, der es ermöglicht, die Aktivität von Proteinen und anderen biologischen Makromolekülen ferngesteuert zu regulieren und bildet somit die Grundlage für die Regulierung verschiedener inter- und intrazellulärer Prozesse. Allosterische Proteine erlauben, dass Veränderungen an ihrem aktiven Zentrum durch das Binden eines Liganden an einer zweiten, oft räumlich entfernten Bindungsstelle—dem allosterischen Zentrum—induziert werden.

Trotz der Allgegenwärtigkeit und der fundamentalen Bedeutung der Allosterie in Proteinen ist das allgemeine physikalische Prinzip, das der effizienten und präzisen langreichweitigen Weiterleitung mechanischer Signale zu Grunde liegt, noch immer ein Rätsel. Die Entdeckung eines allgemeinen Mechanismus würde nicht nur zum Verständnis grundlegender Prozesse beitragen, die für die Aufrechterhaltung des Lebens entscheidend sind, sondern birgt auch ein signifikantes praktisches Potenzial, ermöglicht es doch die Vorhersage von allosterischen Zentren in Proteinen und damit die Adressierung des derzeitigen Engpasses bei der Entwicklung allosterischer Arzneimittel.

Aufbauend auf dem Erfolg, den spektrale Methoden für das Verständnis der Proteindynamik bisher gezeigt haben, stellen wir hier eine Hypothese für ein einfaches physikalisches Prinzip vor, das allosterisches Verhalten in aus Proteinen generierten und in trainierten, künstlichen elastischen Netzwerken einheitlich erklärt. Das Prinzip beruht auf dem Konzept eines kollektiven Hebels, der die Kopplung zwischen steifen und flexiblen Moden koordiniert. Dabei spannen die durch Binden eines Liganden induzierten Deformationen kollektive steife Federn um das allosterische Zentrum herum an. Die Reaktion darauf hingegen verläuft entlang weicher Moden, welche somit die gespeicherte Energie effizient in nicht-lokale Deformationen am aktiven Zentrum umwandeln.

Um diese Hypothese zu überprüfen, entwickeln wir ein schnelles und genaues Verfahren zur Bestimmung der vollständigen (nichtlinearen) Antwort elastischer Netzwerke auf Störungen, die durch Deformationen an den allosterischen Zentren verursacht werden, und verwenden es um Netzwerke evolutionär so zu trainieren, dass sie allosterische Antworten aufweisen. Wir beobachten nichtlineares und nicht-reziprokes Verhalten in den allosterischen Antworten und diskutieren dessen Ursprung sowie die Auswirkungen, die daraus für Allosterie im Allgemeinen folgen mögen.

Mit spektralen und störungstheoretischen Ansätzen finden wir überzeugende Belege und bestätigen die vorhergesagten Charakteristika, die die Besonderheit der echten Bindungstasche im Vergleich zu anderen möglichen Bindungsstellen in von Proteinen abgeleiteten und künstlich trainierten Netzwerken ausmachen sollen.

Schließlich demonstrieren wir die Anwendbarkeit des Prinzips für die Entwicklung von Arzneimitteln, indem wir es zur Vorhersage allosterischer Zentren in Proteinen einsetzen—mit bemerkenswertem Erfolg.

Diese Erkenntnisse verschieben das Paradigma der allosterischen Signalausbreitung in Netzwerken und Proteinen von einer rein auf weichen Moden basierenden Interpretation hin zu einem zweistufigen Denken, das stark zwischen dem allosterischen In- und Output differenziert. Wir kommen zu dem Schluss, dass die Allosterie nur dann vollständig begreifbar ist, wenn sie als das betrachtet wird, was sie tatsächlich ist—ein Nicht-Gleichgewichtseffekt, bei dem die Störung in der spektralen Interpretation explizit berücksichtigt werden muss.

Schlagwörter *Allosterie, Proteine, Konformationsänderungen, elastisches Netzwerk, kollektive Moden, Modenkopplung, Nichtlinearität, Nichtreziprozität, evolutionäre Optimierung, Netzwerkdesign, Arzneimittelentwicklung*

Contents

Chapter 1

Introduction

1.1 Second secret of life

A single living cell contains thousands of proteins that catalyze chemical reactions. Multiple of these reactions occur in close proximity, both temporal and spatial. The product of one reaction often resembles the substrate of another, such that a complex network of metabolic pathways emerges [Alberts et al., 2020]. Evolution equipped the cell with sophisticated tools allowing for precise control of these networks, enabling it to respond to changes in its environment and thus maintain homeostasis. Aside from compartmentalization and other ways of varying the local concentration of proteins, the most general and widespread method of regulating these reactions takes place in form of control loops. The principle is depicted in Fig. 1.1. Such feedback inhibition (or activation) manifests itself either indirectly via modulation of the gene expression of the protein associated with the regulation reaction, or directly by modification of the corresponding protein in the pathway itself.

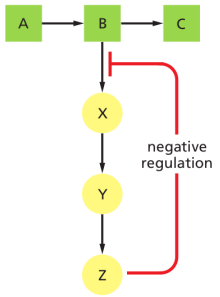


Figure 1.1: Feedback inhibition of a biochemical reaction pathway.

Allostery, termed the *second secret of life* [1] [Fenton, 2008], provides the underlying basis for this by rendering proteins molecular switches. Allosteric proteins exhibit a change (not necessarily a conformational one) at their active site in response to binding a ligand at another, often spatially distant site—the allosteric site. Allosteric inhibition stands in direct contrast to competitive inhi-

[1] Directly following the genetic code.

bition, where the inhibitor binds to the same site as the substrate does. This fact was also key for coining the term allostery, as it describes a form of inhibition *where the inhibitor is not a steric analogue of the substrate* [Changeux, 2013]. The two ancient Greek words that are combined to form *allostery* [Monod and Jacob, 1961], ἄλλος (*allos*) στερεός (*stereos*), translate to *other* and *object*.

Allosteric effects, however, are neither limited to occur in proteins alone nor do they control only biochemical reactions. Macromolecules such as of DeoxyriboNucleic Acids (DNAs) [Garcia et al., 2012, Kim et al., 2013], RiboNucleic Acids (RNAs) [Peselis et al., 2015, Peselis and Serganov, 2021] and complexes of protein and RNA [Walker et al., 2020, Williams and Hall, 2014] are also known to exhibit allostery. Allostery is furthermore an important mechanism utilized in signaling [Bu and Callaway, 2011], e g. neurotransmitter receptors such as ligand-gated ion channels undergo allosterically controlled conformational changes in order to initiate action potentials along neurons [Changeux, 1966, Fealey and Hinderliter, 2013, Neuroscience, 2008].

1.2 Molecular switches

The first evidence for a switch-like behavior in cells came in the form of measurements of Christian Bohr [Bohr et al., 1904], showing that the oxygen-haemoglobin binding curve was not, as previously proposed [Hüfner, 1890], hyperbolic but instead a sigmoidal shape, demonstrating positive cooperativity, as shown in Fig. 1.2 (a). The regulatory role of cooperation in metabolic pathways was first correctly assigned to the effect in 1954 [Boell et al., 1954]. As similar observations accumulated [Umbarger, 1956, Yates and Pardee, 1956] the need for an explanation of the new class of *feedback-inhibited enzymes* became stronger.

At this time, there was neither structural information of the enzymes nor biochemical insight about the binding sites available. For this reason, the interpretation of the *action at a distance* phenomenon, shown in Fig. 1.2 (b) deserves considerable appreciation [Motlagh et al., 2014].

1.3 Early phenomenological models

Structure determination using x-ray crystallography allowed access to protein structures in both conformations, free and ligand bound [Perutz et al., 1960]. This initiated the development of phenomenological models which have proven successful in quantify-

ing the thermodynamics and kinetics of allostery in various proteins [Changeux, 2012]. Two different conceptual themes have been established early on and continued to influence the field until today [Ravasio et al., 2019]. The so called *population shift* or Monod-Wyman-Changeux [Monod et al., 1965] and *induced fit* or Koshland-Némethy-Filmer [Koshland Jr et al., 1966] models both assume the protein to exist in two distinct conformational states, tense or relaxed, with a higher binding affinity for the substrate in the latter. They differ mainly in their assumption of the pre-accessibility of the two states. As the name suggests, both states are accessible simultaneously in the *population shift* mechanism and it is their relative population that is changed upon binding of the fist ligand. In contrast, there is only one state accessible in the *induced fit* model and it is the ligand binding that induces the conformational change of state, thereby modifying the affinity for the other ligand. Despite their differences, Eigen developed a general scheme that covers both models [Eigen, 1968].

These early models of allostery were strongly influenced by the first available crystallographic structures of proteins, in particular the overall quite symmetric tetramer hemoglobin [Perutz et al., 1960]. Over time the concepts have developed further, along

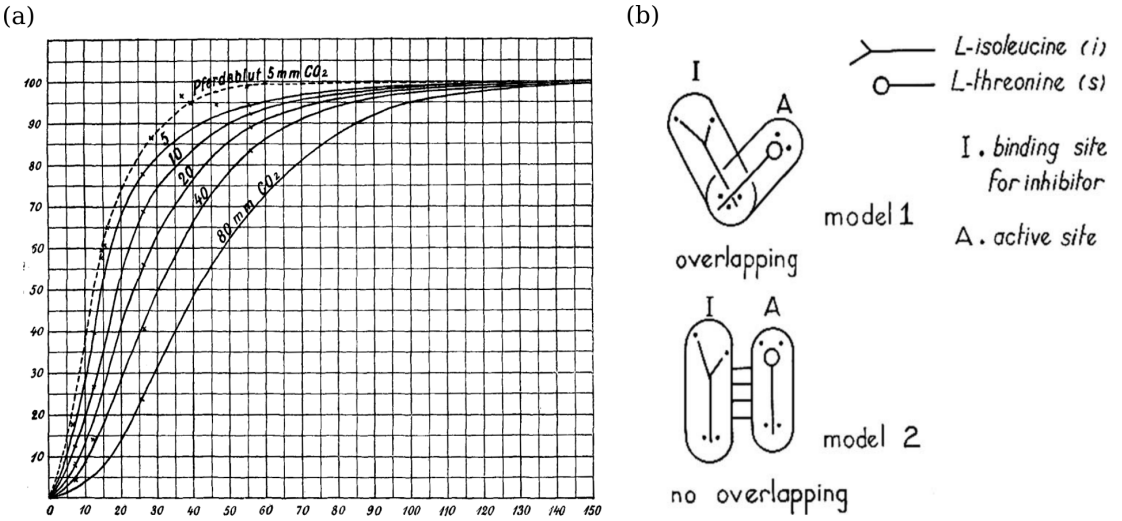


Figure 1.2: On the left (a) the sigmoidal binding curves of haemoglobin that initiated the quest for understanding allostery. The curve shows the partial pressure of oxygen in blood (abscissa) vs. the fraction of saturated, oxygen-bound hemoglobin. On the right (b) the original drawing depicting the two models considered to explain feedback inhibition in L-threonine deaminase (showing similar sigmoidal binding curves). Model 1 refers to competitive inhibition at the orthosteric site which could not explain the shape described by the data, while model 2, later termed allosteric, assumed coupling between distinct binding sites and was in agreement with the data. Figures taken from [Bohr et al., 1904, Changeux, 2011], respectively.

with the improvement of experimental techniques. The focus shifted away from static structures alone [Perutz, 1970] and more towards the role of dynamics of proteins, especially with the advancing deployment of Nuclear Magnetic Resonance (NMR) methods [Wand, 2013]. While first only multimeric proteins were thought to display allostery, it was later also found in monomeric proteins [Ascenzi and Fasano, 2010].

Numerous conceptual models have been proposed, ranging from landscape models incorporating multiple conformational states [Cuendet et al., 2016] to models that address allosteric effects even in intrinsically disordered proteins [Hilser et al., 2012], or in cases where there is no apparent conformational change [Motlagh et al., 2014].

Despite their broad and successful application, phenomenological models have their shortcomings. The underlying reason for the structural coupling between the allosteric and functional sites cannot be resolved with a purely thermodynamic approach [Cui and Karplus, 2008, Kar et al., 2010, Thirumalai et al., 2019]. Even though a conformational change is inherently linked to the *population shift* and *induced fit* concepts, a quantitative connection is still elusive [Tsai and Nussinov, 2014].

1.4 Structural insights

Insights into the allosteric mechanisms of *individual* proteins can been gained combining experimental, computational and theoretical methods. The following overview is not exhaustive. Computational methods mainly employ physics based Molecular Dynamics (MD) simulations with up to atomistic resolution to gain insight into allosteric communication [Laine et al., 2012, Weinkam et al., 2013, Dixit and Verkhivker, 2011, Smith et al., 2016], an approach that is of course still limited by computational power to sub-microsecond timescales [Hospital et al., 2015]. Evolutionary analysis [Rodriguez et al., 2010] aims to find conserved residues that hint at important allosteric pathways and falls into the realm of computational biology. Theoretical approaches cover graph theory [Amor et al., 2016, del Sol et al., 2006, Kaya et al., 2013], distance geometry [Greener et al., 2017] and statistical structural analysis [Atilgan et al., 2007, Mitternacht and Berezovsky, 2011, Tee et al., 2018].

Crystallographic and NMR experiments are able to decipher the various different mechanisms of allostery in proteins [Selvaratnam et al., 2011]. For example, covariance analysis of NMR chemical shifts revealed networks of coupled residues [Laskowski et al., 2009]. By means of an analysis the strain differences between distinct crystal structures of proteins

in the ligand bound and unbound states identified a network of residues connecting the allosteric and active sites [Mitchell et al., 2016].

With growing numbers of biomolecular structures resolved with high resolution in different conformations, the main method to describe allosteric transitions was to compare these and to define the allosteric motion as the conformational change upon binding of the ligand at the allosteric site. Systematic studies of these transitions found that most of the motion happens in less constrained regions of the proteins with allosteric proteins undergoing far more pronounced motions than non-allosteric proteins [Daily and Gray, 2007].

So far a plethora of different allosteric inputs has been disovered, ranging from binding of small molecules or other proteins over phosphorylation to disulfide bond modification. The responses are not less diverse, covering the opening, closing, or other rearrangements of the active site, changes in rigidity or electrostatic properties of the active site, or the modification of overall dynamic properties [Laskowski et al., 2009].

1.5 Pharmaceutical relevance

The fact that enzymes can be controlled remotely, that is via their allosteric site, is of great practical importance in pharmacological applications. Traditionally, drugs are designed to target the orthosteric pocket; allosteric drug design on the contrary optimizes ligands for binding to the allosteric pocket.

This offers multiple advantages, starting with an increased potential to modulate proteins that have proven to be undruggable at their active site [Chatzigoulas and Cournia, 2021]. The non-competitive binding of allosteric drugs allows a stronger modulation effect at smaller concentrations due to the sigmoidal shape of the binding curve [Guarnera and Berezovsky, 2019]. As a result of a lower evolutionary pressure [Kar et al., 2010], the allosteric binding site may be more species-specific [Nussinov and Tsai, 2013], allowing to selectively address subtypes within families of otherwise similar receptors [Wenthur et al., 2014] and leading to far less off-target effects, thus reducing or preventing side effects and drug toxicity [Guarnera and Berezovsky, 2016]. Interactions between proteins are especially challenging to modulate with drugs, a limitation that is lifted when targeting a potentially existing allosteric site [Laskowski et al., 2009].

The main limiting factor in designing new allosteric therapeutics is the knowledge of allosteric binding pockets. High-Throughput Screening (HTS) of an extensive library of drug candidates for a given desired effect is frequently employed to provide ini-

tial chemical leads in the early phases of pharmaceutical discovery [2]; it is so far the main driver for the identification of both the binding sites and the corresponding ligands [Laskowski et al., 2009, Ni et al., 2021a].

A long list of other experimental techniques has been developed, but despite their strengths, these techniques all have the major disadvantage of being time-consuming and expensive. [Tee et al., 2018] In contrast, computer assisted search is much more economical, but has its own shortcomings. Sequence-based techniques try to infer allosteric sites from co-evolving amino acids, obtained via Multiple Sequence Alignment (MSA) of homologous sequences [Teşileanu et al., 2015]. Their predictive power is strongly limited by the variability of the interactions between ligand and binding site even for proteins that have similar sequences. Deducing binding pockets directly from the structure is another strategy, as is the analysis of topological and/or physicochemical properties that leads to numerous drug candidates [Tian et al., 2018, Huang et al., 2013].

Considering that most, if not all [Gunasekaran et al., 2004] proteins display allostery to a certain degree, allosteric drug design opens a promising avenue, especially for proteins that are not druggable at their active site. Understanding the mechanisms underlying allosteric signal propagation can therefore represent a great support for discovering allosteric sites in proteins and thus aid drug design.

1.6 Elusive general principle

As we have seen, allostery is of enormous interest from both, a basic scientific and a practical perspective. There is, on the one hand, the deep desire to unravel the fundamental processes that are important for sustaining life, and on the other hand the tremendous impact allostery implies for pharmaceutical applications.

However, although allostery is ubiquitously found in proteins and is exceedingly well understood in the case of a plethora of individual examples, the physical basis underlying mechanical signal propagation is not clear. This is exemplified by review articles in the field repeatedly throughout the last decade, stating e. g.

> *Although allosteric regulation is the 'second secret of life', the molecular mechanisms that give rise to allostery currently elude understanding.* [Fenton, 2008]

[2] An overview of recently approved drugs found via this route are given in [Macarron et al., 2011]

> *In spite of individual successes in understanding the structural determinants of allostery in well-documented systems, much less success has been achieved in identifying a set of quantitative and transferable ground rules that provide an understanding of how allostery works.* [Hilser et al., 2012]

> *In spite of its importance, allosteric mechanisms in most instances remain a biophysical enigma, eluding a general, quantifiable and predictive atomic description.* [Motlagh et al., 2014]

A fact that contributes to the difficulty of generalizing allosteric mechanisms is that, although some allosteric responses are—with admittedly a significant portion of imagination—classifiable into human understandable motions, e. g. *hinge, shear, piston-like, twist, rocking* or combinations of these, many proteins resist such a classification [Gerstein and Krebs, 1998, Taylor et al., 2014]. The reason for this is simply the nonexistence of discrete categories; there is a continuum of different mechanisms available for proteins to choose from [Liu, 2021].

In recent years the trend has been to view allostery as a property that arises rather generally in mechanical systems, which can be described in terms of networks of interactions between the individual elements [Dokholyan, 2016]. For proteins, the model routinely used for this purpose is the Elastic Network Model (ENM), a single-parameter network model which describes interactions between amino acids with Hookean springs [Bahar et al., 2010]. In this case, a local perturbation at one site of the network (e. g. , binding of a ligand at the allosteric pocket) is then propagated through the network and elicits a specific response at another site. The specific arrangement of amino acids in the protein provides for a sophisticated interplay of interactions which channels the perturbation, normally decaying as power law in elastic media [Landau and Lifshitz, 1986], instead over long distances to the desired target.

This assumption is supported, first of all, by the fact that allostery is found in all combinations of biomolecules, not limited to proteins. Moreover, effects very similar to allostery have very recently been shown in artificial networks (both two- and three-dimensional), which are highly similar to ENMs [Yan et al., 2017, Rocks et al., 2017, Flechsig, 2017, Kim et al., 2019, Kim et al., 2022].

This similarity of effects in a wide variety of systems suggests a common, deeper origin of allostery. The successful extension of allosteric concepts to (one-dimensional) flow networks has affirmed this emerging trend to view allostery as a property of complex networks in general [Rocks et al., 2020, Rocks et al., 2021].

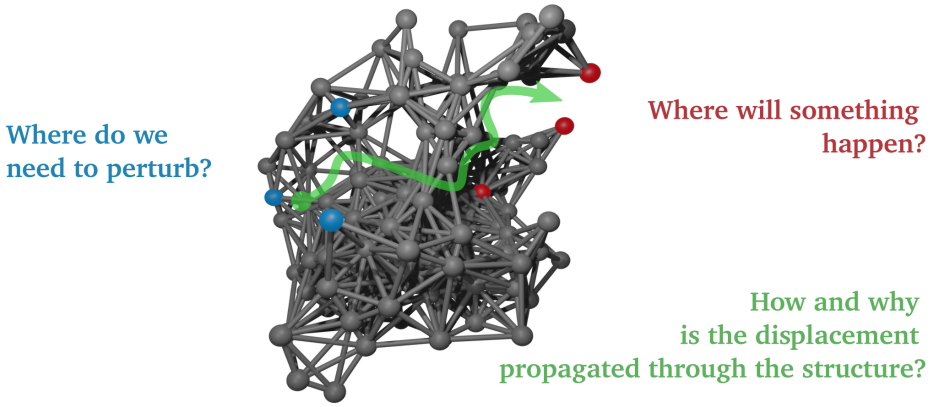


Figure 1.3: A schematic depicting the set of open questions that we address in this thesis.

1.7 Scope and outline of the thesis

After we devoted this chapter so far to the relevance of understanding allostery and acknowledged the need for a unifying principle, we will now outline the structure of this work and formulate the questions we aim to answer.

Building upon the success of network descriptions in the understanding of protein motion, and the ease with which artificial networks could be trained to show behavior that is remarkably similar to allostery, we enqueue in the row of physicists trying to address complex biological problems using minimal physics based-models.

This work aims to answer the following two main questions:

- What is the general physical principle underlying the propagation of displacements in mechanical allosteric networks? Is it the same for protein-derived and artificial networks and therefore responsible for the observed similarities?
- With drug-design applications design in mind, is there a difference between allosteric source and target sites? Can we use this knowledge to predict possible allosteric pockets?

These questions are visualized in Fig. 1.3. The structure of the thesis is as follows: We review the theories and concepts important for describing the dynamics of Elastic Network Models (ENMs) in **Chapter 2**. The reason for the equivalent or even superior performance of Normal Mode Analysis (NMA) with ENMs over traditional force fields is discussed; the

difference between soft and stiff eigenmodes and the current interpretation of their function is pointed out; possible methods for computing the effect of a binding perturbation on networks are compared. As we will show there is controversy in the field in terms of whether or not the response of proteins to the perturbation induced by ligand binding can be described using linear approximations. The current state of designing artificial allosteric networks is reviewed at the end of **Chapter 2**.

Based on physical intuition, in **Chapter 3** we formulate a simple hypothesis that specifies the general principle underlying the large-scale mechanical signal propagation in allosteric structures. Unlike a simple physical lever, a collective lever based on the interplay of soft and stiff modes is supposed to govern allostery.

In **Chapter 4** we briefly discuss the structures that we analyze, in particular we describe how we train a set of 30 ENMs to display a specific (generally nonlinear) response, and how we determine possible candidates for binding pockets in both, the artificial networks and 14 allosteric protein-derived networks. Analytic expressions for the correlations between distances in source and target pockets of trained and untrained networks are derived and shown to be insufficient to distinguish between allosteric and random networks in equilibrium.

In **Chapter 5** we develop a method to accurately evaluate the full, i. e. nonlinear response to ligand binding events in a fast and simple manner. This algorithm is used to train our own set of allosteric networks which we in turn analyze with respect to their ability to transport mechanical information from their source to target sites. We observe nonlinearity and nonreciprocity during the responses and discuss the implications this may have for allostery.

We thoroughly test the novel hypothesis in **Chapter 6** and find it convincingly confirmed for artificial (i. e. trained) and protein-derived networks. Applying the proposed principle, we are able to predict allosteric source pockets with remarkable success, considering that only the initial structure and a single-parameter model is employed. Our approach is therefore suitable to serve as a screening tool in allosteric drug design, by ranking possible binding sites which are to be then assessed with greater chemical detail with any of the much more time and resource consuming computational HTS methods using e. g. MD simulations.

Finally in **Chapter 7** we conclude with a short perspective and provide an outlook towards possible extensions.

Chapter 2

Background

2.1 Harmonic approximation

Problems involving the interaction of more than two bodies yield equations that are in general not solvable analytically. Some famous examples, where an exact solution can be derived even for systems with more than two bodies, are the ideal gas, the harmonic crystal, and the two-dimensional Ising model. While nowadays computer simulations are applied broadly to determine solutions to Newton's equations of motion for molecular systems with many interacting particles, a lot can (still) be learned from approximations that were employed before such numerical methods were available.

In systems, where deviations from the stable equilibrium are small enough, it is possible to describe motions in terms of coupled harmonic oscillators. Assuming time-independent, conservative systems, where the potential energy only depends on the particles' positions, a system is considered to be in a mechanical equilibrium if there are zero forces acting on it. Using the coordinates $\boldsymbol{r} = \{r_i\}$, where i denotes both the particle and the cartesian indices, we can write

$$\left.\frac{\partial U}{\partial r_i}\right|_{\boldsymbol{r}^{(0)}} = 0 \qquad \forall_i, \tag{2.1}$$

indicating the equilibrium configuration with $\boldsymbol{r}^{(0)}$. If the equilibrium is a (local) minimum of the potential function, it represents a (locally) stable configuration. Small enough deviations $\delta\boldsymbol{r} = \boldsymbol{r} - \boldsymbol{r}^{(0)}$ from this initial configuration can be conveniently described using

an expansion of the potential U. Expanding the potential in a Taylor series around the equilibrium positions $\boldsymbol{r}^{(0)}$ up to the second order, we obtain

$$U(\boldsymbol{r}) \approx U(\boldsymbol{r}^{(0)}) + \sum_k \left.\frac{\partial U}{\partial r_k}\right|_{\boldsymbol{r}^{(0)}} \delta r_k + \frac{1}{2}\sum_k \sum_l \left.\frac{\partial^2 U}{\partial r_k \partial r_l}\right|_{\boldsymbol{r}^{(0)}} \delta r_k \delta r_l. \tag{2.2}$$

Setting the arbitrary zero of the potential to be equal to the equilibrium potential, the constant term can be dropped. Due to the equilibrium condition in equation 2.1 the linear term in equation 2.2 vanishes, too. We are left with a pure quadratic form,

$$U(\delta\boldsymbol{r}) = \frac{1}{2}\delta\boldsymbol{r}^{\mathsf{T}}\underline{\mathrm{H}}\delta\boldsymbol{r}, \tag{2.3}$$

where we introduced the Hessian matrix $\underline{\mathrm{H}} \in \mathbb{R}^{n\times n}$ as shorthand for the combined second order derivatives. $\underline{\mathrm{H}}$ is symmetric and positive semi-definite. A similar expansion can be made for the kinetic energy, leading to

$$T(\delta\dot{\boldsymbol{r}}) = \delta\dot{\boldsymbol{r}}^{\mathsf{T}}\underline{\mathrm{M}}\delta\dot{\boldsymbol{r}}, \tag{2.4}$$

with $\underline{\mathrm{M}}$ being the mass matrix containing the particle masses m_i and $\delta\dot{\boldsymbol{r}}$ denotes the time derivative of the displacement.

This leads to the Lagrangian

$$L = \frac{1}{2}(\delta\dot{\boldsymbol{r}}^{\mathsf{T}}\underline{\mathrm{M}}\delta\dot{\boldsymbol{r}} + \delta\boldsymbol{r}^{\mathsf{T}}\underline{\mathrm{H}}\delta\boldsymbol{r}). \tag{2.5}$$

The corresponding equations of motion read

$$\underline{\mathrm{M}}\delta\ddot{\boldsymbol{r}} + \underline{\mathrm{H}}\delta\boldsymbol{r} = 0, \tag{2.6}$$

which is a system of n harmonic oscillators, coupled by the off-diagonal terms of the matrices $\underline{\mathrm{M}}$ and $\underline{\mathrm{H}}$. The general solution for this system is a linear combination of the general solutions of the individual oscillator equations. However, $\underline{\mathrm{M}}$ and $\underline{\mathrm{H}}$ are typically not diagonal. The linear transformation that renders both matrices simultaneously diagonal returns new coordinates, eigenmodes, which we will treat in the next section.

We must keep in mind however, that this approximation allows to describe motions that occur within the vicinity of this initial minimum only. Describing complex motions originating from potentials exhibiting more structure requires sophisticated techniques that typically

involve numerical approaches. Depending on how strong the deviation from the first minimum is, partially analytic methods can still be appropriate, as we will see in Chapter 5.

2.2 Spectral decomposition

Using an exponential Ansatz $\delta\boldsymbol{r} = \boldsymbol{a}e^{-i\omega t},\ \omega \in \mathbb{R}$, the coupled differential equations in Eq. (2.6) can be turned into a linear system of equations. The decoupling works via solving the corresponding general eigenvalue problem

$$(\lambda_\nu \underline{\mathrm{M}} - \underline{\mathrm{H}}) \cdot \boldsymbol{a}_\nu = 0 \qquad \forall_\nu, \tag{2.7}$$

with $\lambda_\nu = \omega_\nu^2$. The transformation matrix $\underline{\mathrm{V}}$ simultaneously diagonalizes $\underline{\mathrm{H}}$ and $\underline{\mathrm{M}}$,

$$\underline{\mathrm{V}}^\mathsf{T}\underline{\mathrm{M}}\,\underline{\mathrm{V}} = \mathbb{1} \qquad \wedge \qquad \underline{\mathrm{V}}^\mathsf{T}\underline{\mathrm{H}}\,\underline{\mathrm{V}} = \underline{\Lambda}.$$

$\underline{\Lambda}$ is diagonal and contains the n eigenvalues λ_ν. The columns of $\underline{\mathrm{V}}$ contain the eigenvectors or eigenmodes of the system, linear combinations of which represent the general solutions to Eq. (2.6):

$$\delta\boldsymbol{r}\,(t) = \sum_\nu \boldsymbol{a}_\nu \Re\{C_\nu e^{-i\omega_\nu t}\}, \qquad C_\nu \in \mathbb{C}, \tag{2.8}$$

with C_ν following from the initial conditions of the system.

The transformation between the displacement $\delta\boldsymbol{r}$ in cartesian coordinates and normal coordinates $\boldsymbol{v}$, written in the basis of $\boldsymbol{a}$, follows

$$\delta\boldsymbol{r}\,(t) = \underline{\mathrm{V}}\boldsymbol{v}\,(t). \tag{2.9}$$

This allows us to write the energy in normal coordinates:

$$U(\boldsymbol{v}) = \boldsymbol{v}^\mathsf{T}\underline{\Lambda}\boldsymbol{v} = \sum_\nu \lambda_\nu \boldsymbol{v}_\nu^2. \tag{2.10}$$

Each normal mode corresponds to the collective oscillatory motion of particles with exactly one of the n eigenfrequencies ω_ν, the inherent resonant frequencies of the system. These correspond to the curvature of the potential energy along the direction of the associated eigenvector. Depending on the dimensionality of the system, a certain number of the eigenvalues is always equal to zero. For tridimensional systems this number is six, corresponding to the degrees of freedom of a rigid body, of which there are three for the translation and

three for the rotation. A zero eigenvalue means there is no restoring force for the motion of this eigenmode, indicating a flat potential along said direction. Mathematically this represents the condition of $\underline{\mathrm{V}}$ being only positive *semi*-definite. The set of normal modes is orthogonal and forms a complete basis in the n dimensional vector space.

If we embed the system in an aqueous environment, we need to account for the effect this has on the dynamics. Motions inside fluids, on the one hand, experience drag, which dissipates kinetic energy into the heat bath, and, on the other hand, undergo small and rapid velocity changes caused by thermal fluctuations. For small (bio-)molecules this behavior is known as *Brownian motion*, first observed by [Brown, 1828] and repeatedly shown to be an appropriate description for proteins as well [Fenimore et al., 2002, Fenimore et al., 2004, Hong et al., 2012]. The two effects, the viscous drag $\underline{\Gamma}\delta\dot{\boldsymbol{r}}$, with the friction coefficient matrix $\underline{\Gamma}$, and the erratic motion have the same origin: $\boldsymbol{\xi}(t)$, the random forces of the small solvent molecules that constantly bombard the observed molecule. The thermodynamic consistency relation between both is given by a Fluctuation-Dissipation Theorem (FDT), which reads

$$\langle \xi_i(t)\xi_j(t')\rangle_{\mathrm{eq}} = 2k_{\mathrm{B}}T\,\underline{\Gamma}_{ij}\delta(t-t'), \tag{2.11}$$

and was first described by [Einstein, 1905]. The forces are assumed to be uncorrelated, stationary and Gaussian distributed. The effect of the solvents can thus be modeled implicitly via a stochastic and a viscous term in the otherwise explicit description of the system, resulting in a *Langevin Equation* [Gardiner et al., 1985]:

$$\underline{\mathrm{M}}\ddot{\delta\boldsymbol{r}}(t) = -\underline{\mathrm{H}}\delta\boldsymbol{r} - \underline{\Gamma}\dot{\boldsymbol{r}}(t) + \boldsymbol{\xi}(t). \tag{2.12}$$

Assuming the dynamics occurs along a path that lies in a single energy valley and no barriers are crossed—an assumption we will thoroughly test and validate—we can accurately describe the motion in terms of its thermal average $\langle\delta\boldsymbol{r}\rangle_{\mathrm{eq}}$ and average out the random force $\langle\boldsymbol{\xi}(t)\rangle_{\mathrm{eq}} = 0$. For the sake of brevity we switch the notation accordingly from here on and refer by $\delta\boldsymbol{r}$ to its thermal average.

Another well-justified assumption in aqueous media is to drop the inertial term $\underline{\mathrm{M}}\delta\ddot{\boldsymbol{r}} = 0$, as we are interested in timescales $t \gg m_i/\gamma_i$, where the momentum is completely dissipated. We obtain the overdamped equation for coupled harmonic oscillators:

$$\gamma\delta\dot{\boldsymbol{r}} + \underline{\mathrm{H}}\delta\boldsymbol{r} = 0, \tag{2.13}$$

which describes dynamics falling into the regime of a *low Reynolds number* [Purcell, 1977]. The assumption of overdamped dynamics is not only valid concerning the medium which surrounds the protein; the intramolecular dynamics is governed by the same physics, as the protein interior is fluid-like[McCammon et al., 1977] with an internal friction comparable to water [Ansari et al., 1992, Ansari, 1999].

This equation is solved by the same diagonalization as used for equation 2.6, the difference being in the exponential Ansatz which is real in this case, $\delta\boldsymbol{r}\,(t) = \boldsymbol{a}e^{\lambda t},\ \lambda \in \mathbb{R}$, such that $1/\lambda$ represents a relaxation time instead of a squared frequency.

2.3 Normal modes in proteins

The aforementioned harmonic approximation and subsequent spectral decomposition is fruitfully applied to study the motion of biomolecules. The method is called Normal Mode Analysis (NMA) in this context and, although in principle agnostic with respect to the function describing the potential energy, in classical NMA one employs the semiempirical Molecular Dynamics (MD) potentials [Levitt, 1983, Weiner et al., 1984, Brooks et al., 1983], the analytic expression for which can be found in the Appendix, Eq. (A.1). The routines for NMA have for quite some time been included in state of the art MD software [Van Der Spoel et al., 2005, Hess et al., 2008, Pronk et al., 2013, Abraham et al., 2015], often different potentials and solvent models are available.

The second ingredient of NMA is the initial structure of the molecule of interest. In the case of proteins it is determined either by means of X-ray crystallography [Perutz et al., 1960], NMR [Güntert, 2009] or cryo Electron Microscopy (cryo-EM) [Wu and Lander, 2020]. The harmonic approximation (see Section 2.1) which NMA is based on demands this initial structure to be at equilibrium, a condition which is generally not satisfied for potentials used in MD simulations. Thus, in order to apply the formalism, an energy minimization step is required, which is computationally demanding[1] and can distort the original structure [Cui and Bahar, 2005].

However, as we will see, these distortions are small and do not strongly impart what we can learn from NMA. The pairs of normal modes and accompanying frequencies describe the motion of the atoms in a structure on different scales. Normal modes belonging to low frequencies (small eigenvalues) describe highly collective, large-scale global motions. The

[1] Essentially, machine precision has to be reached for the eigenvalues to become non-negative, which is computationally much more expensive than the energy minimization that renders structures stable enough for MD simulations. The calculation of the Hessian matrix is performed via finite differences of the force function, resulting in additional computational effort.

high frequency modes describe more localized motions, which tend to be smaller but faster, like the oscillations of side-chains or other small groups of atoms. Often one keeps the bond lengths fixed in classic NMA, allowing only for rotations around the joints [Go et al., 1983]. As the characteristics of the slow modes does not change with this approximation, it is assumed that these oscillations involving large groups of atoms are governed by averaged interactions and are independent of microscopic (aka chemical) details. This is supported by the fact that, when the density of the spectrum is studied for different proteins, the curves happen to collapse onto a universal curve [Ben-Avraham, 1993], indicating that there must be the same physics governing this large scale behavior.

Quite early it has been observed that an albeit in this case phenomenologically found and thus an imposed single global mode can bridge two enzyme conformations with and without a ligand bound, which indicated that global modes are relevant for function [McCammon et al., 1976]. The connection to allostery can already be drawn here as it was not the ligand that was bound to the catalytic site of the enzyme but actually the inhibitor at the allosteric regulatory site. A hinge mode similar the the one reported in [McCammon et al., 1976] could also be found using NMA in [Marques and Sanejouand, 1995]. Low frequency motions, although not always assignable to some human understandable mechanism, have been found to account for up to 95% of the atomic displacements in proteins, when compared to full MD simulations using the same potential [Levy et al., 1982]. An observation, that has ever since been confirmed numerous times [Noguti and Gō, 1982, Go et al., 1983, Levitt et al., 1983] even when NMA is performed in vacuum [Brooks and Karplus, 1983].

Note here that of course proteins do not exist in vacuum but are embedded in a solvent that is mainly composed of water and ions. The expected overdamped and diffusive motions are observed [Kitao et al., 1991] and the *Langevin modes* [Lamm and Szabo, 1986] recovered. The respective study of eigenmodes in the overdamped setting is also termed *Brownian mode analysis* [Hinsen et al., 2000].

Therefore, as anticipated in the last section, we must let go of the idea of harmonic periodic oscillations around the equilibrium structure and think more in terms of repeated, solvent-induced perturbations away from—and subsequent returns toward—the equilibrium with a decay rate that is given by the eigenvalues [Togashi et al., 2010]. That of course means that the ligand induced conformational motions must be described in an overdamped setting as well. Solvent related effects, like the hydrophobicity of proteins, can be taken into account to a certain degree using implicit solvent models in the MD potentials [Lazaridis and Karplus, 1999].

Collective modes can also be extracted from trajectories of MD simulations using e. g. Principal Component Analysis (PCA). They are not limited to a single potential energy well, directly take solvent effects into account, and due to the sampling nature incorporate entropy directly. Of course, MD trajectories need to be simulated before these principal modes are accessible, which comes with a major drawback; a large portion of timescales relevant for proteins (> ms) are out of reach even for highly optimized, parallel algorithms running on state-of-the-art specialized hardware [Shaw et al., 2010]. Reassuringly however, the soft modes from NMA often overlap substantially with the collective modes from PCA of MD simulations [Hayward et al., 1993].

One popular early application of NMA was the refinement of crystallographic structures via their B-factors which are directly proportional to the atomic mean square displacements [Diamond, 1990, Kidera and Gō, 1992].

Thus, notwithstanding the caveat that the movements are assumed to occur only in the initial minimum, a lot can be deduced from there.

2.4 Elastic network models

The Elastic Network Model (ENM) is a simple, physics based model aimed to describe the long-timescale dynamics of proteins on a coarse-grained level. The networks' nodes are beads and the edges Hookean springs connecting the beads. Depending on the level of coarse-graining, the beads can represent different chemical entities, ranging from atoms, as used in the original publication [Tirion, 1996], over amino acids [Haliloglu et al., 1997] up to capsomers of viruses [Tama and Brooks III, 2005]. In principle the method is agnostic with respect to length and time scales.[2]

The energy of a system of N beads is given by summing the energies of all of the pairs that are connected by a spring:

$$U(\boldsymbol{r}) = \frac{1}{2} \sum_{i<j}^{N} \underline{\mathrm{A}}_{ij}\, \kappa_{ij}\, (r_{ij} - r_{ij}^{(0)})^2. \tag{2.14}$$

Here $\boldsymbol{r}$ represents the positions of the beads, and the distance between two beads i and j is $r_{ij} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}$, with the superscript (0) indicating the initial inter-bead distance. κ is the spring constant, often assumed to be identical for all con-

[2] Note that coarse-grained NMA on the amino acid level is not possible with the traditional potentials [Hayward and Groot, 2008].

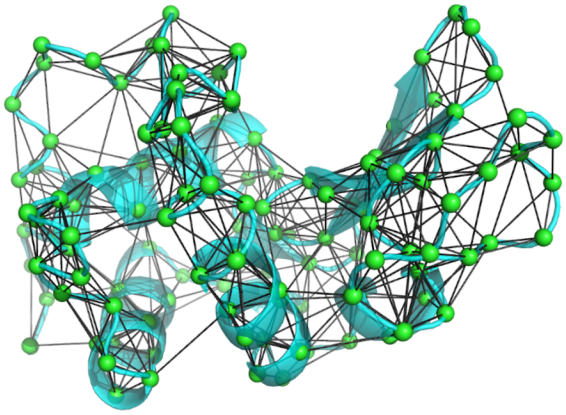


Figure 2.1: A cartoon representation of a protein; overlayed is the ENM with springs depicted by lines which connect amino acids closer than a certain cutoff. Figure taken from [López-Blanco et al., 2014]

nections. The adjacency (or connectivity) matrix $\underline{\mathrm{A}}$ represents the underlying graph and encodes which beads are connected,

$$\underline{\mathrm{A}}_{ij} = \begin{cases} 1 \text{ for } r_{ij}^{(0)} \leqslant r_c \\ 0 \text{ for } r_{ij}^{(0)} > r_c \end{cases} . \tag{2.15}$$

Despite the mathematical similarities to the early network models of Flory [Flory, 1944] and Rouse [Rouse, 1953] the ENM of Tirion is conceptually different. While both types of models employ springs to connect the chemical entities (nicely termed *submolecules*), the harmonic potential in the earlier models is meant to describe the actual interaction between the chemically or physically bonded units, while the Hookean potential in Tirion's version needs to be understood in an effective manner. As the early network models are meant to describe flexible linear chains of polymers, they do not insert springs between beads that are not bonded. Although Flory later extended his ideas to more general network structures still only the cross-links, i. e. physical connections of polymers occur in the connectivity matrix [Flory, 1976]. Flory and Rouse both neglect excluded volume effects and hydrodynamic interactions. The springs have a zero equilibrium length, only the effect of the heat bath the system is immersed in prevents the positions to coincide. Conversely, Tirion's model assumes all beads with an initial inter bead distance below a certain cutoff length r_c to be connected by harmonic springs with a *non-zero equilibrium length*. Thereby even long range forces like electrostatics are taken into account to a certain degree. The forces obviously are not valid for arbitrary constellations of the beads, but are meant as restoring forces towards the initial structure. This restriction to describe motions

only around a specific structure essentially includes excluded volume implicitly. The old network models for flexible polymers are not limited to a single configuration.

It was conjectured in earlier studies using classical NMA that the large collective motions of proteins are independent of the chemical details [Ben-Avraham, 1993]. Based on this assumption, Tirion replaced the sophisticated MD potentials with a number of springs that represent the effective potential that she believed emerges from a combination of numerous individual interactions. The specific form of these interactions is rendered irrelevant at larger scales due to the mixing and averaging of multiple interactions. Tirion even relates the specific form of the harmonic potential to the central limit theorem.

ENMs feature the beautiful property of giving access to collective motions of proteins from their structure alone without much computational effort.[3] Although initially meant to replace the traditionally employed potentials for NMA, they are now fruitfully deployed in a wider range of setups, sometimes even replacing the standard potential in MD simulations, e. g. when studying the effect of ligand binding [Flechsig, 2017], nonlinear responses to perturbations [Togashi and Flechsig, 2018] or in a pulling setup for unfolding proteins [Poma et al., 2018].

Two different classes of ENMs exist: the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM). They differ in their dimensionality, for a network of N beads a GNM is a N dimensional, while a ANM is a $3N$ dimensional model. The former just deals with the amplitudes of motions while the latter also encodes directions. In the GNM the Hessian matrix is reduced to an N dimensional matrix, closely related to the Kirchhoff matrix from graph theory. We will deal only with ANMs in this work as we are *per se* interested in the directionality of motions; we will further perform a coarse-graining on the residue level.

2.5 Normal modes of elastic networks

As indicated in section 2.3, ENMs were first introduced to the world of proteins by Tirion [Tirion, 1996], in order to replace the complicated, energy minimization requiring, semiempirical potentials which were conventionally employed for NMA [Marques and Sanejouand, 1995].

[3] This holds of course for other structures as well; the method is routinely applied in engineering. Instead of computationally demanding finite-element analyses, a so called *lumped mass-spring* model gives a good initial approximation of the dynamics [Tian et al., 2010].

Following directly from the way ENMs are set up, i. e. all springs are inserted relaxed, at their equilibrium length, it can be anticipated that no energy minimization is required. The Hessian matrix $\underline{\mathrm{H}}$ can be evaluated analytically for this simple potential, its off-diagonal super-elements read

$$\underline{\mathrm{H}}_{kl} = \left(\frac{1}{r_{kl}}\right)^2 \begin{pmatrix} x_{kl}^2 & x_{kl}y_{kl} & x_{kl}z_{kl} \\ y_{kl}x_{kl} & y_{kl}^2 & y_{kl}z_{kl} \\ z_{kl}x_{kl} & z_{kl}y_{kl} & z_{kl}^2 \end{pmatrix} \Bigg|_{\mathrm{r}=\mathrm{r}^{(0)}}, \tag{2.16}$$

with r_{kl} denoting the absolute distance between beads k and l, and x_{kl}, y_{kl}, z_{kl} the respective entries along the three cartesian axes individually. The diagonal super-elements follow

$$\underline{\mathrm{H}}_{kk} = -\sum_{l=1, l\neq k}^{N} \underline{\mathrm{H}}_{kl}. \tag{2.17}$$

Thus, only the diagonalization of $\underline{\mathrm{H}}$ requires numerical linear algebra.

In addition to the technical advantages in calculating ENMs, their low parameter requirements and simple analytical functions have methodological advantages as well. First, a reduction of fitting parameters always yields a stronger predictive power of a model. To underscore that point: The single free parameter in ENMs is the cutoff length r_c.[4] Second, the fact that it is only the combination of a high number of interactions and not the individual details that govern the large scale dynamics of proteins (around their current configuration) allows for an easily understandable interpretation.

A typical example of a system exhibiting a conformational motion that is accessible by ENM is the mechanism of force generation in the motor protein myosin, which has been explained using NMA of ENM in [Zheng and Brooks, 2005a]. Even events usually thought to be inaccessible to ENMs, such as protein unfolding, can be studied with NMA, and the sequence and regions where local unfolding occurs could be predicted successfully [Su et al., 2008].

Solvent effects are hard to account for, but in principle, an implicit solvent extension of the Hessian matrix is possible thanks to tedious work, where the second derivatives of the Generalized Born framework have been evaluated analytically [Brown and Case, 2006]. The beauty of the simple ENM would however be lost with such an extension, both from a com-

[4] And indeed, even this parameter can be varied over a wide range of values without qualitatively changing the dynamics of the structure studied. In the literature values between 7 and 16 Å are reported [Sanejouand, 2013, Kondrashov et al., 2007, Kundu et al., 2002].

putational and systematical point of view. The analytic expressions, filling pages, rely on symbolic algebra programs and constructing the implicit solvent extended Hessian matrix takes substantially longer than the standard one in Eq. (2.16) [Brown and Case, 2006].

ENMs are frequently applied to refine NMR [Gniewek et al., 2012] and cryo-EM [Zheng, 2011, Tekpinar, 2018] structures. The number of applications of ENMs is growing, a wealth of examples can be found in the following reviews: [Eyal et al., 2006, Lezon et al., 2009, Sanejouand, 2013, López-Blanco et al., 2014, Togashi and Flechsig, 2018].

The computational effort for NMA with ENM nowadays is negligible. At the end of the last century it was possible to obtain normal modes of a system with about 50 residues in *a few minutes on standard desktop computers* [Hinsen, 1998], for large systems containing several hundreds to thousands of particles for a long time the limiting factor was memory [Hayward and Groot, 2008]. Today an equivalent calculation is performed on the order of a second and memory is rarely a problem. Even the diagonalization of extremely large matrices for systems containing several hundred thousand particles is feasible within minutes when optimized algorithms and hardware are employed [Lopez-Blanco et al., 2013]. With the increasing amount of automated webservers that are available (two examples are [Bakan et al., 2011, Suhre and Sanejouand, 2004]) the input-output ratio of effort to information is so low, that there are even voices proposing ENMs should become a *routine part of a structural investigation* [Bauer et al., 2019] for biologists.

Soft modes

As Tirion found, the low frequency modes of the Hessian matrix of the elastic network agree almost perfectly with those obtained by classical NMA. This initial finding was soon after supported by a study considering more than 300 proteins [Bahar and Jernigan, 1997]. The assumption of Tirion thus was well justified, and her model supports and extends the idea that similar physics governs soft vibrational modes of macromolecules [Ben-Avraham, 1993].[5] Other comparative studies have shown that these soft, functional modes of proteins are conserved when different versions of ENMs where used to obtain them [Nicolay and Sanejouand, 2006, Na et al., 2018]. The soft, functionally relevant modes exhibited great robustness against mutations [Zheng et al., 2006].

[5] Remarkably, although proteins are almost as densely packed as crystals [Liang and Dill, 2001], their vibrational spectrum differs systematically [Ben-Avraham, 1993, Tirion, 1996] from the one proposed for the latter[Debye, 1912] in that the density of states scales linearly instead of quadratically with the frequencies.

Large-scale collective conformational motions induced by ligand binding often show significant overlap with the soft modes in a protein. The overlap however is not always greatest with the softest mode, large overlaps within the first few tens of modes are not uncommon [Zheng and Doniach, 2003]. For simple motions like the often reported hinge-mode, these overlaps tend to be much higher when compared with the modes in the open structure rather than the closed one [Tama and Sanejouand, 2001]. When a conformational motion spans multiple potential wells, the soft modes are able to describe the motion at least at the beginning of a conformational change [Togashi et al., 2010].

Fig. 2.2 depicts an example of soft modes in a protein in the upper row.

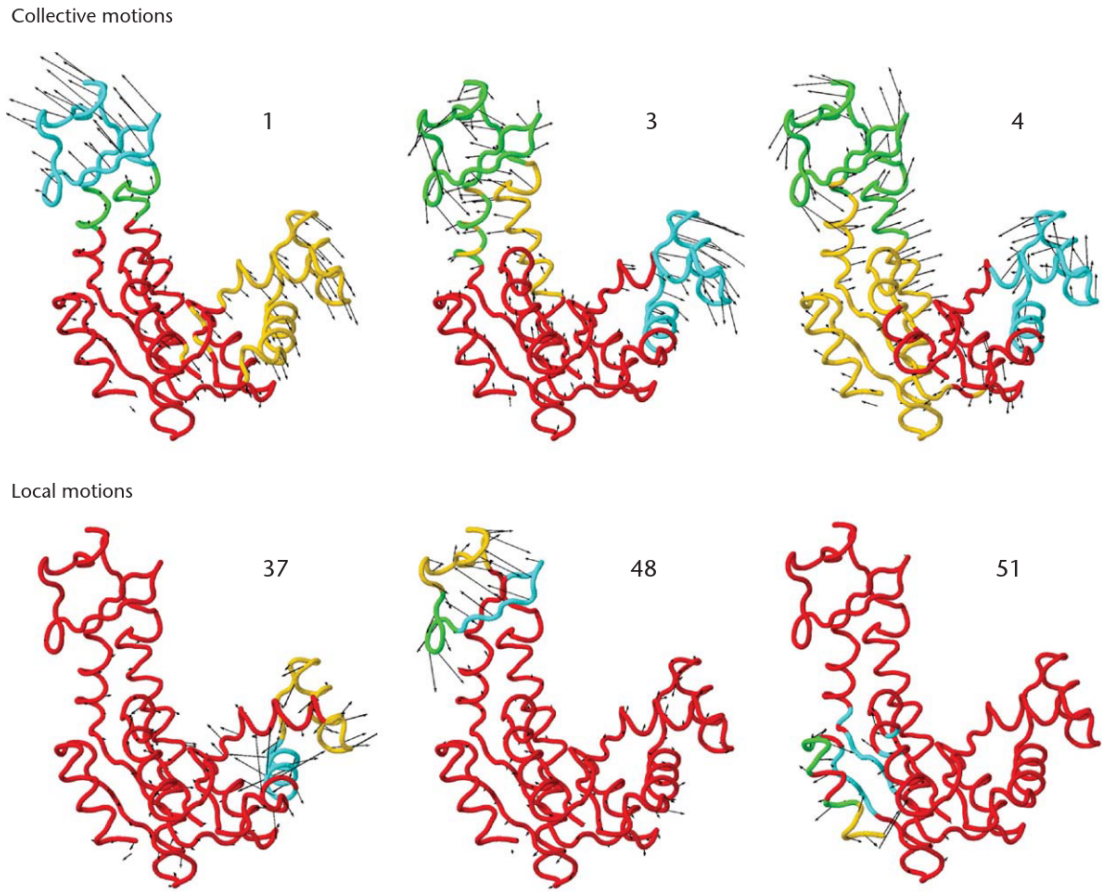


Figure 2.2: The difference between soft, global (upper row) and stiff, local (lower row) modes in an ENM of a protein. Figure taken from [López-Blanco et al., 2014]

Stiff modes

Modes belonging to small eigenvalues are called soft because they require little energy to be excited. In contrast to these are the intermediately high modes[6] that we term stiff as their corresponding entries in the Hessian represent stiffer combinations of springs. In other words, stiff modes belong to higher eigenvalues which encode a steeper curvature of the energy function, meaning they show higher resistance against displacements.

These modes have been far less studied. Nevertheless, they carry nontrivial and valuable information. First of all, stiff modes are much more localized in the networks and thus involve smaller groups of beads [López-Blanco et al., 2014]. The reason for this strong localization can be traced back to inhomogeneities in the structure of proteins and comparable media [Bahar et al., 1998, McLeish et al., 2013]. As anticipated by Anderson, impurities can cause modes to localize within a few wavelengths, which is a general wave phenomenon observed in different systems [Anderson, 1978, Anderson, 1958].

The regions where these stiff modes localize often represent the binding sites of enzymes [Sacquin-Mora et al., 2007], a finding that has been independently confirmed by studies comparing NMR and X-ray structures using PCA [Yang et al., 2009]. In such binding sites the residues are known to be more rigid, showing lower crystallographic [Bartlett et al., 2002, Yuan et al., 2003] and theoretically determined B-factors [Yang and Bahar, 2005]. The rigidity is also found to be tightly related to the local packing density [Halle, 2002].

Fig. 2.2 clearly depicts the difference between soft, large-scale collective modes and stiffer, more localized modes. It is widely accepted that soft modes are relevant for a protein's function due to their frequently observed large overlap with experimentally resolved conformational changes. However, stiff modes are also occasionally suspected to participate in function (e. g. in electron transport proteins) and residues participating in stiff modes are thought to be critical for stabilizing the folded structure [Bahar et al., 1998]. Later in this work, we will expand on their responsibilities in appreciable detail.

2.5.1 Linear response

The framework of Linear Response Theory (LRT) allows the study of systems that are driven (slightly) out of equilibrium by a weak external perturbation using only the knowledge of the fluctuations at equilibrium [De Nittis and Lein, 2017].

[6] When using an all-atom ENM, the very high modes at the upper end of the spectrum require knowledge of the chemical details and thus cannot be accurately described by a simplified low-resolution model [Tirion, 1996].

The setup perfectly lends itself to study the effect of a perturbation, e. g. ligand binding, on the equilibrium fluctuations of the protein which can be understood as normal modes in the unbound state of the protein.

The equilibrium probability[7] written in terms of the eigenmodes and eigenvalues reads

$$P_{\mathrm{eq}}(\boldsymbol{v}) = \frac{1}{Z} \int e^{-\frac{1}{2}\sum_{\nu} \lambda_\nu \boldsymbol{v}_\nu^2} \, d\boldsymbol{v}, \tag{2.18}$$

with the partition function being

$$Z = \int e^{-\beta U(\boldsymbol{v})} \, d\boldsymbol{v} = (2\pi)^{\frac{n}{2}} \prod_{\nu}^{n} \lambda_\nu^{-1/2}. \tag{2.19}$$

Then the fluctuations in equilibrium are given by

$$\langle \boldsymbol{v}_\nu^2 \rangle_{\,\mathrm{eq}} = \frac{1}{Z} \int P_{\mathrm{eq}}(\boldsymbol{v}) \boldsymbol{v}_\nu^2 \, d\boldsymbol{v} = \frac{1}{\lambda_\nu}. \tag{2.20}$$

This follows from solving the generalized Gaussian integral.

The fluctuations in cartesian coordinates thus are simply given by the inverse of the Hessian matrix,

$$\langle \boldsymbol{r}\boldsymbol{r}^{\mathsf{T}} \rangle_{\,\mathrm{eq}} = \underline{\mathrm{H}}^{-1}, \tag{2.21}$$

allowing in turn to determine the response to an external force via

$$\delta\vec{r}_i = \sum_j \langle \vec{r}_i \vec{r}_j^{\mathsf{T}} \rangle_{\,\mathrm{eq}} \, \vec{f}_j. \tag{2.22}$$

A more direct and intuitive approach would be to see the quadratic approximation of the potential energy as a linear approximation of the force, arriving at essentially the same formula:

$$\boldsymbol{f} = \underline{\mathrm{H}} \delta\boldsymbol{r}. \tag{2.23}$$

This framework has been applied one-to-one in numerous studies. First, deriving the fluctuations from both MD simulations and an ANM potential, the conformational changes related to ligand binding determined in crystallographic studies could be very well re-

[7] We are throughout measuring energies in units of $k_{\mathrm{B}}T$ and n is equal to $3N - 6$, such that we are only summing over the indices belonging to nonzero eigenvalues.

produced for a set of three proteins. Interestingly, the authors reported that the response does not agree with a single mode. Indeed, it is a combination of up to five soft modes that is required in order to achieve high overlaps with the perturbation induced response [Ikeguchi et al., 2005].

Proteins show highly anisotropic responses in single-molecule pulling experiments, i. e. the resistance against the pulling force varies strongly depending on which residue the force is applied to. This anisotropy could largely be reproduced in an equivalent setup where the response to forces applied to an ENM was computed with LRT. The difference in the deformations depending on where the forces acted was shown to be caused by the participation of soft modes at the corresponding residues, which are easier to deflect than stiff modes at other residues. However, they also discovered cases where the LRT breaks down, indicating that barriers are crossed, and thus the original energy minimum is left early during the response [Eyal and Bahar, 2008].

In a related study, a tool was developed that allows to scan over all residues of a protein, observing the response that perturbations at different sites have on the whole structure. The allosteric mechanism behind the efficient uptake and release of iron in an ferric binding protein was unraveled this way [Atilgan and Atilgan, 2009].

In all of these studies, the ligand binding was mimicked by forces acting on a subset of the beads. A different formulation also allows to calculate the linear response to closing a binding pocket: Interpreting the response of the elastic network as the new set of equilibrium coordinates given a constraint on the positions of the perturbed beads, the linear response problem can be rephrased as a constrained minimization problem [Rocks et al., 2017]. The minimum is then found demanding that the Lagrangian is extremal with respect to both, the multipliers and the induced displacements. Quite similar to this idea is also the formalism introduced in [Wyart, 2005] and used in [Yan et al., 2017]. The difference only lies in the implicit use of the constraint, i. e. no Lagrangian multipliers are employed here explicitly. The first application of this method to proteins may be found in [Zheng and Doniach, 2003].

2.5.2 Nonlinearities

The aforementioned method reaches the limits of validity when the real response of the network starts to leave the original minimum. This, however can only be detected when the full nonlinear response is available for comparison. In a computational study using the

full (nonlinear [8]) ENM potentials it has been shown that the range where NMA is applicable can vary strongly between systems. Both of the studied motor proteins showed nonlinear behavior in their relaxation to equilibrium after random perturbations were applied. The last part of the relaxation trajectories does agree with a slow mode. However, the part of the nonlinear relaxation trajectory that agrees with the slow modes is negligibly small in one of the examples in [Togashi et al., 2010]. The authors point out that in order to describe the nonlinear relaxation, a single harmonic approximation does not suffice, even if many modes are involved. Curiously, the same authors emphasized three years earlier in a similar study, how remarkably well the linear normal mode description works for such relaxation studies of ENM of proteins and artificial networks [Togashi and Mikhailov, 2007], emphasizing the strong dependence on the protein actually studied.

One possible way to describe nonlinear transitions is to iteratively combine ENMs to provide a path that connects between the two known allosteric end states of a protein. The high dimensional path is then constructed from consecutively choosing linear combinations of up to three soft modes which showed the greatest overlap with the empirical conformational change. A further extension that allows for partial unfolding makes it possible to approach the open conformation of the protein adenylate kinase, starting from the closed structure. The reverse direction is generally simpler to reproduce with ENM. This study confirms that only a part of the allosteric response is accurately described by the initial energy minimum [Miyashita et al., 2003].

Thus, whenever the response occurs in a sequential manner that spans multiple distinct energy minima [Frauenfelder et al., 1991], as is known to be possible for proteins [Daily and Gray, 2007, Formaneck et al., 2006], a nonlinear description is required to fully describe allosteric transitions. We will address these issues in detail in Section 5.4.

2.6 Artificial allosteric networks

2.6.1 Fundamentals

Although ENMs were originally developed to describe proteins they can in fact be used to model any mechanical system in which some underlying network of links imposes constraints on the position of nodes while allowing small fluctuations driven by thermal noise.

[8] The forces of the springs are linear in terms of their beads' pairwise distances. Due to the nonlinearity of the distance function however, the forces are nonlinear with respect to the positions of the beads [Togashi and Flechsig, 2018].

Examples may include nano-machines such as piezoelectric actuators that move probe-tips in atomic force microscopes [Tian et al., 2010, Sierra et al., 2005].

The interest in the motion or rather their suppression in structures strikingly similar to ENMs dates back to Maxwell [Maxwell, 1864]. Although the applied principle was initially derived in an earlier work [Clapeyron and Lamé, 1831], it was the first systematic study of the flexibility of networks. These early studies lay the foundations of the theory of structural rigidity [Crapo, 1979].

A network in three dimensions, back then termed *frame*, was defined as *a system of lines connecting a number of points*. In a stiff or rigid frame, the distance between two points can not be altered without simultaneously changing the length of at least one of the connecting lines.[9] A structure that is not rigid exhibits one or more Inextensional Mechanisms (IMs). Such a mechanism can emerge either trivially, when the network has too few connections, or requires a more sophisticated analysis, e. g. when a particular arrangement of links allows for it. If a redundant link exists in a structure it is said to exhibit a State of Self-Stress (SSS). In special cases, elegantly observable in the tensegrity structures of Buckminster Fuller, these SSS can impart stiffness to multiple IMs [Calladine, 1978].

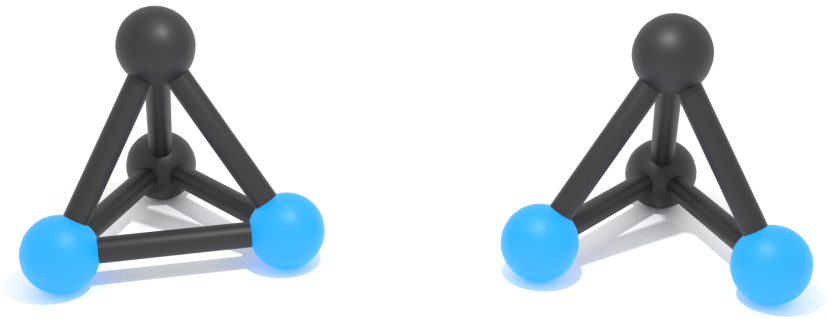


Figure 2.3: A schematic representation of a rigid (left) and nonrigid (right) structure. The unstable structure can display an inextensional mechanism. This alters only the distance between the blue beads, which are not connected with each other. Figure produced by the author and published in [Lapolla et al., 2021].

If j is the number of joints and b the number of links in a three-dimensional structure, Maxwell's well known condition for the stiffness of frames states

$$\mathrm{b} = 3\mathrm{j} - 6, \tag{2.24}$$

[9] Such a rigid network is equivalent to an ENM of a protein and can still perform conformational motions.

with 6 indicating the rigid body motions. The number of rigid body motions in d dimension is $\frac{d(d+1)}{2}$. However, as already anticipated by Maxwell, in more complex settings, the pure topology of the network does not suffice to classify the mechanics in frames and, whoever tries to do so, may *end up in a state of confusion* [Pellegrino and Calladine, 1986]. In general, the full geometrical specification, i. e. the positions of all joints and directions of all connections are needed to find hidden IM and SSS, of which there are m and s, respectively [Tarnai, 1980]. Using linear algebra, Calladine was able to provide a complete mathematical framework for this problem.

Through a Singular Value Decomposition (SVD) of the equilibrium matrix $\underline{\mathrm{E}}$ both, the numbers and directions of the inextensional mechanisms and of the states of self-stress are accessible, respectively [Pellegrino and Calladine, 1986]. The matrix $\underline{\mathrm{E}}$ connects the forces $\boldsymbol{f}$ acting on the nodes with the tensions $\boldsymbol{t}$ in their links, via

$$\underline{\mathrm{E}}\boldsymbol{t} = \boldsymbol{f}. \tag{2.25}$$

The rank r of the matrix $\underline{\mathrm{E}}$ is intimately related to both m and s, as $\mathrm{s} = \mathrm{b}-\mathrm{r}$ and $\mathrm{m} = 3\mathrm{j}-6-\mathrm{r}$. Maxwells rigidity condition is thereby extended to:

$$\mathrm{b} - 3\mathrm{j} + 6 = \mathrm{s} - \mathrm{m}. \tag{2.26}$$

The two structures shown in Fig. 2.3 are the simplest illustration of this relation for the existence vs. the absence of a mechanism in three dimensions. Both structures have $\mathrm{j} = 4$ nodes and $\mathrm{s} = 0$ states of self-stress. The rigid one with $\mathrm{b} = 6$ bars has no inextensional mechanism ($6 - 12 - 6 = 0 - 0$), while the structure with $\mathrm{b} = 5$ bars has exactly $\mathrm{m} = 1$ of these mechanisms ($5-12-6 = 0-1$). The matrix $\underline{\mathrm{E}}$ is a linear map from the b-dimensional space of tensions to the $3\mathrm{j} - 6$-dimensional space of forces or loads.[10] The directionality of the IM and SSS is encoded in the four fundamental vector subspaces associated with the equilibrium matrix $\underline{\mathrm{E}}$ [Calladine, 1978, Pellegrino and Calladine, 1986]. Any $n \times m$ matrix like $\underline{\mathrm{E}}$ has four subspaces related to it: The column space, the row space (which is the columns space of $\underline{\mathrm{E}}^{\mathsf{T}}$), the (left) null space (which is the right null space of $\underline{\mathrm{E}}^{\mathsf{T}}$), and the (right) null space [Strang, 1993]. The connection between the subspaces and the IM and SSS becomes clear in Table 2.1. The column space and row space of $\underline{\mathrm{E}}$ contain base vectors defining the combinations of loads and tensions, respectively, which can be supported by the structure. The vectors describing the loads are actually identical to the eigenvectors of the corresponding Hessian matrix for the same system. Base vectors of the

[10] A related linear transformation is possible in the space of extensions $\mathbf{e}$ and displacements $\mathbf{d}$: $\underline{\mathrm{E}}^{\mathsf{T}}\,\mathbf{d} = \mathbf{e}$.

Table 2.1: The four vector spaces of $\underline{\mathrm{E}}$ can be computed with a SVD and give access to the the states of self-stress, the inextensional mechanism as well as the motions that would require connections to be stretched.

		Vectorspace	dim	Relation
edge space	$\mathbb{R}^{\mathrm{b}}$ $\left\{ \begin{array}{l} \mathcal{R}(\underline{\mathrm{E}}) \\ \mathcal{N}(\underline{\mathrm{E}}) \end{array} \right.$	Rowspace Nullspace	$\mathrm{r}_{\underline{\mathrm{E}}}$ s	supported tensions states of self-stress
node space	$\mathbb{R}^{3\mathrm{j}-6}$ $\left\{ \begin{array}{l} \mathcal{C}(\underline{\mathrm{E}}) \\ \mathcal{N}(\underline{\mathrm{E}}^{\mathsf{T}}) \end{array} \right.$	Columnspace left Nullspace	$\mathrm{r}_{\underline{\mathrm{E}}}$ m	supported forces mechanisms

left nullspace of $\underline{\mathrm{E}}$ represent the forces that the structure could not withstand, which are exactly m IMs. Lastly the (right) null space of $\underline{\mathrm{E}}$ contains the SSSs. This analysis goes one step beyond a standard NMA. The SVD can also be interpreted as a generalization of the eigendecomposition in Chapter 2.2 for non-square matrices:

$$\underline{\mathrm{E}} = \underline{\mathrm{U}}\,\underline{\Sigma}\,\underline{\mathrm{V}}^{\mathsf{T}}. \tag{2.27}$$

Indeed, the left and right singular vectors of $\underline{\mathrm{E}}$, found in the columns of $\underline{\mathrm{U}}$ and $\underline{\mathrm{V}}^{\mathsf{T}}$, are the eigenvectors of $\underline{\mathrm{E}}\,\underline{\mathrm{E}}^{\mathsf{T}}$ and $\underline{\mathrm{E}}^{\mathsf{T}}\underline{\mathrm{E}}$, respectively. The singular values (non-zero diagonal elements of $\boldsymbol{\Sigma}$) are the square roots of the non-zero eigenvalues of $\underline{\mathrm{E}}^{\mathsf{T}}\underline{\mathrm{E}}$.

2.6.2 Bottom-up design

The limit of the characterization of IMs based on linear algebra is that it can detect only the absence of length changes of the connecting edges of first-order, i. e. it cannot distinguish between finite and infinitesimal motions. Going beyond this limitation is the framework developed in [Kim et al., 2019], which allows the full parametrization of the nonlinear shape of IM and culminates in a rigorous design principle for arbitrary motions in frameworks: By fixing the positions and desired motions of a subset of beads and assuming a bipartite network structure (i. e. two groups of beads which are only connected between groups, not within groups) it is possible to solve for the positions and motions of the other sub-

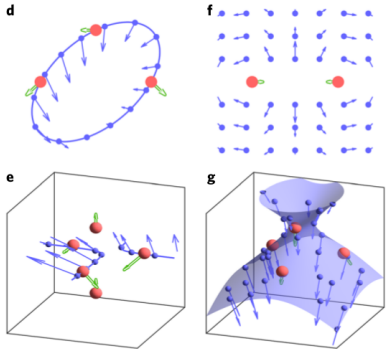


Figure 2.4: One- and two-dimensional solution spaces for IMs, left and right, respectively. Figure taken from [Kim et al., 2019].

set of beads. The solution spaces—these correspond to the IMs but without the limitation of infinitesimal motions only—are then characterized by the edge constraints between the connected beads, and define hyper-surfaces on which the beads of the second set can be placed. These beads and their solution spaces can be seen as blue beads and lines/surfaces, respectively in Fig. 2.4, with the fixed beads and their motions in red and green, respectively.

Incredibly complex nonlinear motions and resulting shapes can be generated with this design principle; compare Fig. 2.5. A crucial ingredient for the principle is, however, the ex-

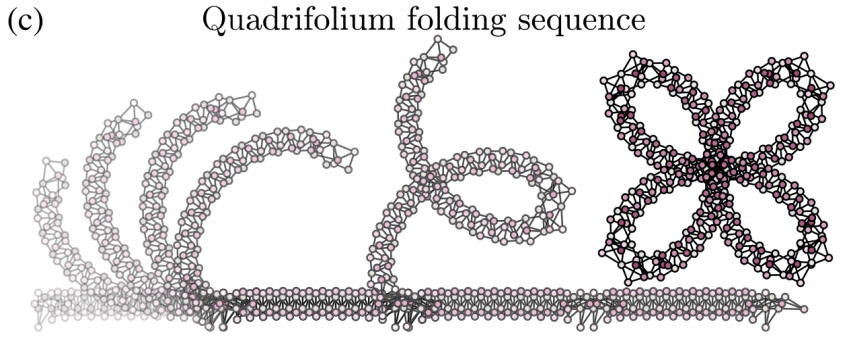


Figure 2.5: Combining many simple four-bar linkages allows to formulate complicated motions using a single parameter. Figure taken from [Kim et al., 2022].

istence of a bipartite graph structure within the framework. That means that some beads, no matter how close their proximity is, must not be connected for the motion to happen, which poses a strong limitation for applying the principle to understand conformational motion in proteins. The effective nature of the physical interactions in ENMs of proteins gives a strong reason for the implicit formulation of the connectivity, which strongly violates the bipartition assumption. Note that in proteins the only IMs are the 6 rigid body motions, the Nullspace of $\underline{\mathrm{E}}^{\mathsf{T}}$, contains no further IM if the cutoff value is chosen correctly [Sanejouand, 2013]. The assumption of completely inextensional linkages is already relaxed in the work itself in order to better describe cooperative effects in elastic networks, the other constraint, an underlying bipartite graph, may possibly be lifted as well, rendering the applicability of the concept to proteins far less stretched. Notwithstanding there is an enormous potential the principle has for applications where one has complete control over the underlying connectivity like, e. g. in the design of robots [Detweiler et al., 2007].

2.6.3 Heuristic design

A completely different route for designing specific behaviors into random spring networks is provided by the heuristic *tunig by pruning* concept introduced in [Goodrich et al., 2015].

First applied in order to tune global properties like the Poisson ratio, it can also be used to incorporate local responses inspired by allostery into the network: A pair of input nodes (source) and output nodes (target) is selected and the structural change at the target site is observed after an applied strain at the source. Calculating the effect that the removal of a single bond has on the strain ratio between the target and source pair, it is possible, by removing only a small set of bonds in the network, to train a long range correlated deformation between the source and target site. The authors report a remarkably high success rate in training the networks, reaching essentially 100% for small strain ratios, but diminishing with higher input strains [Rocks et al., 2017].

Quite similar to this is the training for an allosteric response in another study, where springs are allowed to swap between occupied and unoccupied links in a Monte Carlo (MC) scheme. Two striking observations were made: The response seems to be almost absent in the bulk of the (two-dimensional) networks and reappears at the active (target) site; the architecture connecting source and target site (measured in mean coordination number) shows a trumpet like shape, hinting towards a more rigid lever at the source and a softer, responsive target site. [Yan et al., 2017]. Yan et al. only train two-dimensional networks but assume their findings to be independent of the dimensionality. They base this assumption on the observation that other effects based on microscopic elasticity of amorphous materials are independent of the dimensionality and that this transfers to allosteric behavior [Liu et al., 2011].

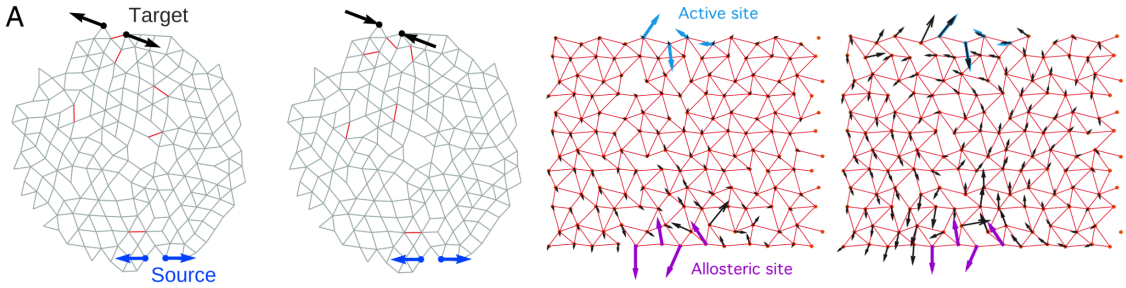


Figure 2.6: Heuristically trained two-dimensional networks. Deleting (right) or swapping (left) of bonds allows to incorporate allosteric behavior into networks with only few changes needed. Figures taken from [Rocks et al., 2017] (left) and [Ravasio, 2020](right).

Both these works employed LRT to calculate the effect of an input displacement on the network structure. Although they consist of beads connected with Hookean springs, the networks in [Rocks et al., 2017, Yan et al., 2017] do not represent ENMs as their connectivity is decided a priori and and only once, thus removing and inserting bonds is straightforward as no cutoff length must be considered.

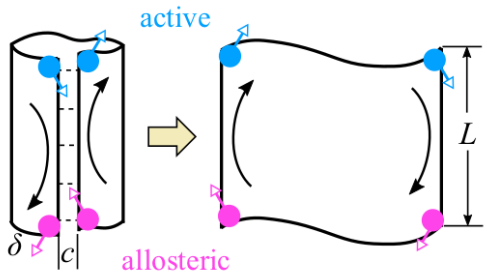


Figure 2.7: Possible emergence of a single soft mode in proteins by embedding structure exhibiting an Inextensional Mechanism (IM) in a soft elastic medium. Figure taken from [Yan et al., 2018].

We concentrate here on networks that perform a single task but the framework is by no means restricted to such scenarios. Multiple independent allosteric responses have been trained in two-dimensional networks already in [Rocks et al., 2017], and the upper limit of simultaneously performable tasks was shown to describe a phase transition, in mechanical and flow networks [Rocks et al., 2019].

In a slightly more biological context, three-dimensional full-featured ENMs are evolutionary trained from random initial configurations to show allosteric responses in [Flechsig, 2017]. The algorithm was already used before to tune soft modes into networks in [Togashi and Mikhailov, 2007]. The beads' equilibrium positions (and thereby also the springs equilibrium lengths) are randomly perturbed in order to change the underlying connectivity of the network, leading to a different response.[11] These studies differ mainly in their fitness function and the number of steps taken in the calculation of the effect of a mutation. [Flechsig, 2017] optimized the full nonlinear allosteric response (several thousand steps required) for a symmetric and an antisymmetric variant originating from the same initial random structure. Conversely, [Togashi and Mikhailov, 2007] optimized several thousand structures in terms of the spectral gap between the first and the second mode—a single step procedure for the response.

An Inextensional Mechanism (IM) as described above in Section 2.6.2 was also discussed as the underlying (presumably, general) principle of allosterically relevant soft modes in proteins. Embedded in a soft elastic material such an IM would obtain a non-zero frequency and could be able to transfer allosteric signals [Yan et al., 2018]. The existence of independent multifunctionality in artificial networks [Rocks et al., 2019] and pro-

[11] As we have adapted the evolutionary training procedure in our work, a more detailed description follows in 5.

teins [Schlessinger, 1986, Light and Anderson, 2013], however, poses great challenges to this explanation.

2.7 Complex networks

Attempts to describe proteins using established theories for other states of matter have so far failed. Although the average packing density in proteins is as high as that in crystalline solids, they rather appear to be liquids or glasses according to their free volume distribution [Liang and Dill, 2001]. Proteins are heterogeneous environments on a single molecule basis but show strong structural and spectral similarities among each other [Ben-Avraham, 1993]. They show percolation behavior that clearly reminds one of a random network [Deb et al., 2009], but display distinct features like soft modes and long ranging responses that differentiate them from the latter.

The existence of a specific long-range mechanical interaction in proteins is also not intuitive at first. Considering that the response to a perturbation decays uniformly and on a rather small lengthscale in media which are similarly densely packed as proteins [Liang and Dill, 2001], the lengthscale over which an allosteric response is propagated is surprisingly large. This reminds of the emergence of floppy modes in fully isostatic systems (which have just enough constraints to maintain rigidity) [Liu et al., 2011], where the lengthscale of elastic responses can reach the system size [Lerner et al., 2014, Düring et al., 2013]. However, proteins can deviate from this critical point and show rigid, isostatic, and flexible regions within a single molecule while still exhibiting long range modes [Buhrow et al., 2012].

The ease with which allostery could be trained into random networks, using different evolutionary schemes and different starting points [Rocks et al., 2017, Yan et al., 2017, Flechsig, 2017] indicates that it must be a property that is somehow inherently ingrained in networks. Almost every studied protein (with a stable fold) exhibits large-scale motions that often overlap with the normal modes either obtained form traditional or ENM based NMA. It is now further proposed that essentially every protein is allosteric [Gunasekaran et al., 2004], and even designing a new allosteric site into a protein was realized [Zhang and Bishop, 2007].[12] Apparently allosteric proteins constitute their own class of systems that is best described by complex networks [Rocks, 2019].

[12] Inspired by allostery in biomolecules, work on implementing allosteric effects in small artificial molecular systems has been successfully carried out in supramolecular chemistry for decades; more than a hundred different mechanisms have been reported so far [Kremer and Lützen, 2013].

We may conclude that the question of how to design allostery has largely been answered. Taking the liberty to employ and extend the available methods, we generate in the upcoming chapters our own allosteric networks and study the general underlying principles of perturbation transmission, which we propose is the physical basis of allostery in proteins.

Chapter 3

Hypothesis

3.1 Towards a mechanism of allostery

Considering that we expect allosteric proteins primarily to exert a switching effect, one can anticipate that there must be a difference between the regulatory (allosteric) site and the controlled (active) site. The mechanical signal should follow a one-way path to prevent the switching effect from being immediately reversed by an action at the controlled site; hence, the mechanism should in general not be reciprocal. More concretely, in the case of elastic networks, we want the allosterically targeted region to be susceptible to motion, and the regulatory site to be particularly good at receiving an input displacement and effectively transmitting it through the network. That is, the source and target pockets should be fundamentally different.

If we focus on the spatial distribution of the different modes in this specific context, we notice that the soft modes appear predominantly at the active site, where the allosteric *response* occurs; the allosteric site itself, however, which accepts the *input*, does not participate in precisely these modes. We will quantitatively confirm this observation in artificial and protein-derived networks in the beginning of Chapter 6.

A related observation is that the allosteric sites are often located in particularly rigid regions, where only the more localized, stiffer modes participate. If we combine these observations with a simple energy conservation argument, we can formulate a general hypothesis for the physical principle underlying the mechanics of allostery.

In order to propagate a mechanical signal as far as possible through an elastic network and translate it into a pronounced motion at another location, the mechanical energy introduced into the network with the signal must be transported in an efficient manner.

The most simple analogy for how this may happen is the classical lever. There, a certain amount of energy is transferred via a short lever d_{in} arm into a large displacement at the end of the long lever arm d_{out}. The relationship between the energy and the lengths of the arms involves the force, which is large at the first location (F_{in}) and small at the second (F_{out}). Simple energy or torque conservation states that

$$F_{\text{in}}\, d_{\text{in}} \;=\; F_{\text{out}}\, d_{\text{out}}. \tag{3.1}$$

If we assume soft springs at the long and stiff springs at the short lever arm, respectively, we can anticipate how the reciprocity vanishes; a fixed input displacement at the short arm can still propagate, while the same input displacement at the long, soft arm would immediately be absorbed by the soft springs.

The networks we are considering are certainly not simple levers. However, a very similar principle is conceivable for those as well.

The local binding of a ligand and the associated closing of the pocket loads the springs located at the source with mechanical energy. Elsewhere, and in particular at the distant target site, where the allosteric effect emerges, this energy is converted into movement. Analogous to the lever, there is also a force in the networks against which work is performed, namely the collective spring constants from the Hessian matrix.

Thinking in terms of the collective normal modes, we remember that the energy can be written as

$$U\left(\delta\boldsymbol{r}(\{\boldsymbol{v}_\nu\})\right) = \sum_\nu \lambda_\nu \boldsymbol{v}_\nu^2. \tag{3.2}$$

The energy associated with an input displacement $\delta\boldsymbol{c}$ written in the diagonal basis reads

$$\delta U(\delta\boldsymbol{c}) = \sum_\nu \lambda_\nu (\boldsymbol{v}_\nu^\mathsf{T}\delta\boldsymbol{c})^2 \tag{3.3}$$

and the corresponding response $\delta\boldsymbol{s}$ gives

$$\delta U(\delta\boldsymbol{s}) = \sum_\nu \lambda_\nu (\boldsymbol{v}_\nu^\mathsf{T}\delta\boldsymbol{s})^2. \tag{3.4}$$

Now it can be easily read off that for a strong transmission of the small and locally confined input displacement, the energy uptake during the input should be as high as possible and the resulting movement during relaxation as large as possible. This would allow for a non-local response, involving larger displacements at the target site when the stored energy is released via the collective soft springs. An allosteric network, therefore, establishes a connection between the seemingly disconnected sites.

The collective lever could be described as conservation of energy related to the excitation of different sets of modes, which are the stiff modes during the perturbation and soft modes during the response:

$$\begin{aligned} \delta U(\delta \boldsymbol{c}) \;&\approx\; \delta U(\delta \boldsymbol{s}) \\ \sum_{\substack{\nu \in \text{stiff} \\ \text{modes}}} \lambda_\nu (\boldsymbol{v}_\nu^{\mathsf{T}} \delta \boldsymbol{c})^2 \;&\approx\; \sum_{\substack{\nu \in \text{soft} \\ \text{modes}}} \lambda_\nu (\boldsymbol{v}_\nu^{\mathsf{T}} \delta \boldsymbol{s})^2. \end{aligned} \tag{3.5}$$

It becomes much clearer for the extremal case where only the stiffest mode is excited by the input displacement and the softest mode relaxes the energy completely[1], then

$$\lambda_{\text{stiff}}\, \boldsymbol{v}_{\text{stiff}}^2 \;=\; \lambda_{\text{soft}}\, \boldsymbol{v}_{\text{soft}}^2. \tag{3.6}$$

Now, as $\lambda_{\text{soft}} \ll \lambda_{\text{stiff}}$ it follows that $\boldsymbol{v}_{\text{soft}} \gg \boldsymbol{v}_{\text{stiff}}$.

Finally, Eq. (3.5) allows us to predict two generic observables that are expected to show distinct behavior for the allosteric coupling in the networks and proteins. First, the projections of the input displacement onto the eigenmodes should be small for soft modes, and the projections of the response onto the modes should be high for soft modes and vice versa. Second, for the real allosteric source beads, the energy uptake during the perturbation *and* the energy release during the relaxation should be largest in magnitude as compared to the other possible pockets.

We will now embark on testing this hypothesis in the remaining part of the thesis.

[1] Although in this case the coupling between the modes in the resting network is indeed impossible due to their orthogonality [Goldstein et al., 2002]; thus we consult the extreme case only to indicate the relation between the eigenvalues.

Chapter 4

Allosteric Structures

4.1 Protein-derived networks

4.1.1 Setup and parametrization

Building ENM of proteins is straightforward. For each protein of interest we retrieve the corresponding file containing the positions and atomic assignments from the Protein Data Bank (PDB) and extract the α-carbons.[1] Then a cutoff value r_c is chosen, assuring that not more than exactly 6 eigenvalues of the Hessian matrix are zero. Springs with stiffness $\kappa = 1$[2] are inserted in between beads which have pairwise distances smaller than the cutoff distance. Whenever the corresponding entry in the PDB contains information about the active and allosteric binding sites this information is used to designate the beads accordingly, otherwise additional publications containing this information are taken into consideration.

4.1.2 Binding-pocket candidates

As a starting point a method is required that allows to find possible binding pockets. This is a thoroughly investigated topic in structural biology, so we decided to adopt an existing method that is acknowledged in the field and easily accessible.

[1] When characterizing a protein composed of a chain of amino acids, often one approximates the position of the amino acids by the position of their respective α-carbon; it is the backbone carbon where different substituents attach to.

[2] This simplification does not affect the qualitative picture of the results and an extension to different values is straightforward.

The algorithm proposed in [Tian et al., 2018] is based on a purely geometric approach, no chemical details enter the procedure of predicting pockets. The basis of the algorithm is the *rolling ball* algorithm [Shrake and Rupley, 1973] frequently used to define the Solvent Accessible Surface (SAS) also called *Connolly Surface* [Connolly, 1983]. The surface is estimated with a probe of the size of the solvent which samples the molecule along the respective van der Waals radii of the atoms. This is essentially equivalent to rolling a ball along the surface. A Delaunay triangulation[3] using the atoms positions is then compared with the solvent accessible surface [Edelsbrunner and Mücke, 1994] and allows to define the molecules' pockets using discrete flows [Edelsbrunner, 1993], a method to discriminate inner and outer cells of the Voronoi diagram [Liang et al., 1998a, Liang et al., 1998b]. It is publicly available through a web interface and returns a list that contains for each pocket found all the residues of the protein that participate.

We find that some of the larger pockets predicted by the algorithm are huge, encompassing up to half of the proteins, see table A.1. To account for the numerous possible ways a ligand could bind within these, we draw combinations of beads out of all the beads in the pocket. These combinations consist of three beads, and per pocket up to a maximum of one thousand are drawn randomly. For the total number of ligand-binding combinations refer to Table A.1.

An example of binding pockets predicted by the webserver is shown in Fig. 4.1. The left panel depicts the volumes of the pockets and the right one the corresponding residues in the ENM.

4.2 Trained artificial allosteric networks

4.2.1 "Growing" pseudo proteins

The structures subsequently trained to display allosteric responses are first "grown" to resemble folded coarse-grained proteins. After the first bead is positioned at the origin, the following steps are iterated until the desired number of beads is reached. Each subsequent bead is placed on a sphere around the previous one with a fixed distance $d_{\min}$, assuring that the distance to all other beads is also above that cutoff value. Two other constraints have to be satisfied. First, all beads must lie within a large sphere of radius d_{big} such that a globular shape is approached. Second, the volume enclosed by two smaller spheres of

[3] In simple terms, a Delaunay triangulation [Delanuay, 1934] is the most efficient way to partition the convex hull of a set of points into triangles or tetrahedra, for two and three-dimensional system, respectively.

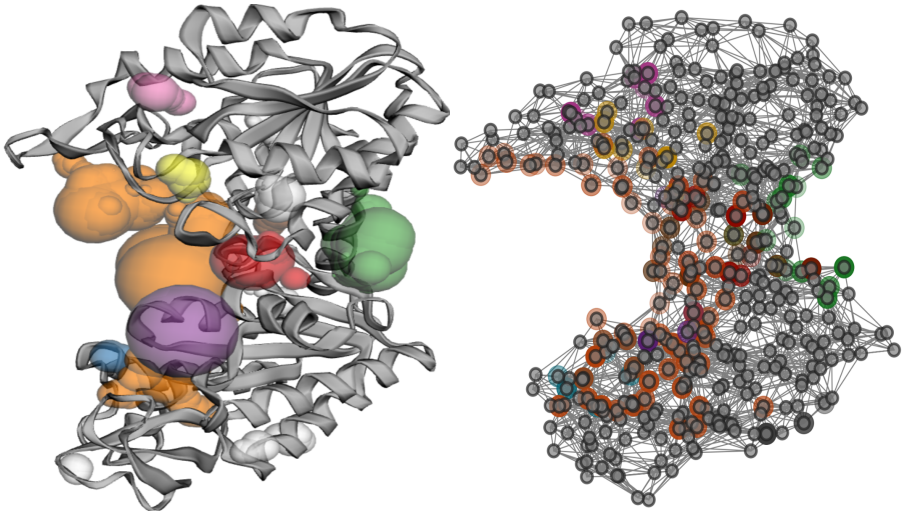
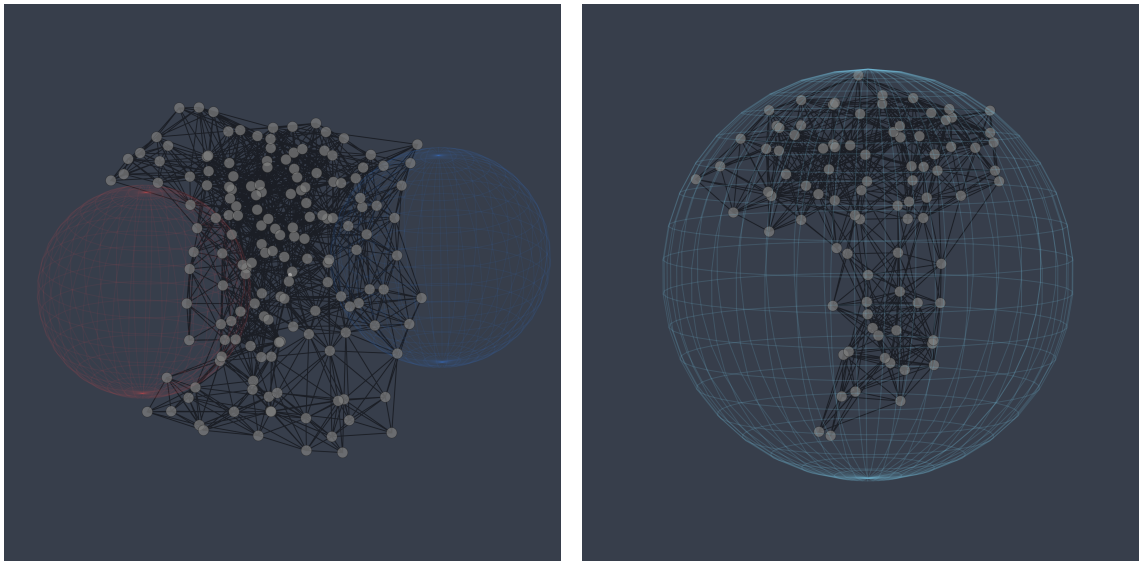


Figure 4.1: Predicted binding-pocket candidates in the proteins are determined geometrically, based on an algorithm from [Tian et al., 2018]. The 9 largest out of 77 total pockets shown. (Left) Cartoon representation of Human Serum Albumin (HSA, PDB ID 2BXD [Ghuman et al., 2005]) as an example protein. (Right) The beads (α-carbons) constituting these pockets are colored accordingly in the ENM.

radius d_{small} which overlap the large one is not populated by any bead, such that pockets on both sides of the sphere emerge. The second condition is relaxed for a subset of structures. For another subset of networks we combine two of the grown structures by merging them and removing beads that are closer than d_{min}. Precise values are given in Table A.2. Manual inspection is required to select the beads that represent the target and source pocket, respectively. The source and target pockets are represented by combinations of two or three beads.

4.2.2 Evolutionary training for an allosteric response

The grown structures are subjected to a training in which a long-range allosteric effect is incorporated. The allosteric effect that is aimed for is specified as follows. It consists of an input and an output; a ligand-binding event that closes the source pocket is mimicked by pulling the source pockets' beads towards their local center of mass. This is what we refer to as *the input*. *The output* refers to the opening or closing of the target binding pocket, depending on the variant being trained. The formalism to compute the full nonlinear response to closing the source pocket and thereby also the closing of the target pocket is discussed in detail in Section 5.1, and resembles the observable that is optimized during the training. We describe the two observables belonging to the pocket closing with their



(a) Pocket spheres **(b)** Outer sphere

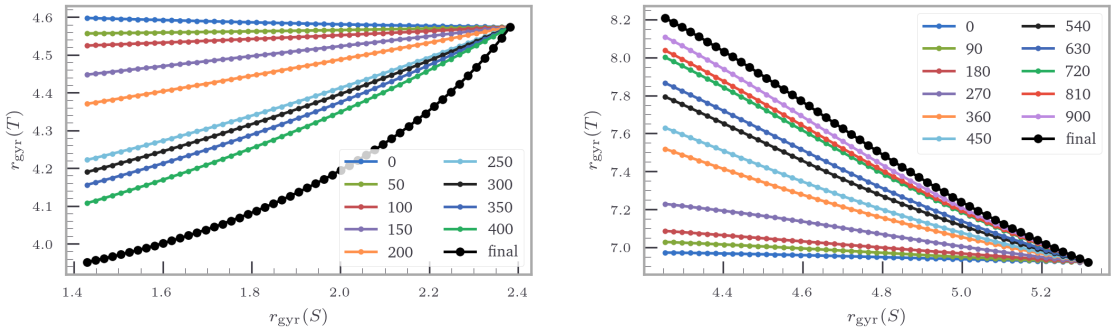
Figure 4.2: Growing scheme slightly modified from [Flechsig, 2017]. The structures are designed to resemble artificial proteins.

respective radii of gyration defined by $r^2_{\mathrm{gyr}}(A) = \frac{1}{N_A}\sum_{i\in A} r_i^2$, where $A = T$ for the target and $A = S$ for the source pocket.

The evolution of the networks happens in an open-system Monte Carlo (MC) fashion, that is, beads can change their position and new beads can be created or existing ones removed. This algorithm is inspired by the natural principle of evolution, corresponding to mutations that change amino acids at a given position in the sequence, or more drastic insertions [King and Jukes, 1969] and deletions [Kimura et al., 1968], respectively.[4] First, one of the three possible steps (move, delete, insert) is randomly chosen with probabilities $(1/2,\ 1/4,\ 1/4)$[5] respectively and the response of the mutated network calculated. If the response at the target site has improved according to the variant trained, the mutation is kept with probability one, otherwise with probability $\exp(-|\Delta r_{\mathrm{gyr}}(T)|)$, where $\Delta r_{\mathrm{gyr}}(T)|) = r^{\mathrm{init}}_{\mathrm{gyr}}(T) - r^{\mathrm{final}}_{\mathrm{gyr}}(T)$ is the change in pocket size during the response. The exponent prevents too big steps backwards while allowing for the circumvention of local minima. Less than a thousand MC steps were required to incorporate the desired effect into the networks.

[4]Depending on the number of deletions and insertions on the DNA level these can cause either frameshift mutations [Bock et al., 2019] (which would change all subsequent amino acids, something we do not account for in our training scheme) or introduce new amino acids [Alberts et al., 2020].

[5]This decision is based on empirical evidence.



(a) Symmetric variant **(b)** Antisymmetric variant

Figure 4.3: Increase in the relative change of the radii of gyration of source and target pocket during the training for two examples of different variants. The radius of gyration of the target changes while the one of the source stays the same.

We provide two examples for a symmetrically and asymmetrically trained network in Fig. 4.3. Upon closing the source pocket a symmetric network closes its target pocket and an asymmetric network responds by opening the target pocket. Both networks are trained from the same initial structure and show essentially zero allostery at the beginning of the training.

4.2.3 Novelty and delineation from previous work

Many methods are available for creating and training allosteric networks. Most closely related to biology (at least to our judgment) is the one described by Flechsig [Flechsig, 2017]. We adopted the method but adapted and optimized it where we saw the need. Briefly summarizing the differences, most of the parameters are re-parametrized and the training of monomeric networks is entirely new. The original publication considered only networks consisting of two of the above described units. Moreover the formulation and evaluation of the response is different and considerably faster. The extension to open MC is also new and deletion and insertion were not accounted for previously.

To remove a possible bias we also considered networks derived from a random “dense packed-spheres” algorithm published by [Baranau and Tallarek, 2017]. These are indicated in Table A.2. This yields structures that are closer to those in [Yan et al., 2017, Rocks et al., 2017]. However, we subjected these structures to the same evolutionary training, as simple flipping or pruning of bonds is not possible within the standard ENM framework following Tirion.

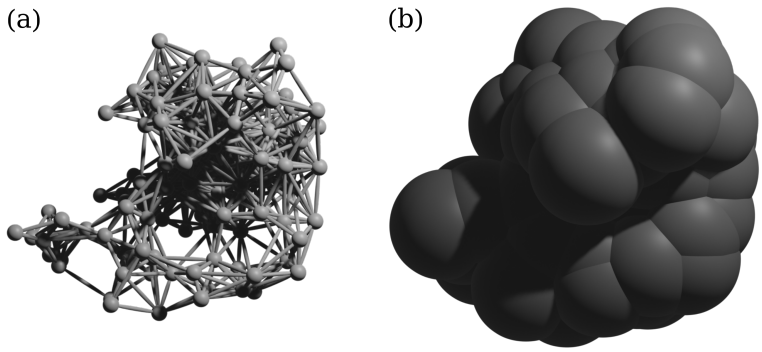


Figure 4.4: We define the surface beads of a network (a) as those beads that contribute to the shared surface after growing the radii of the beads (b). Potential binding pockets are selected from this set of beads.

4.2.4 Determination of pockets on the surface

In particular, in the context of drug design, where we aim to find the unknown allosteric source pocket, a means to find binding pocket candidates is required. We tried to come up with such a method for the artificial networks, where two or three beads resemble a simple approximation of a simple binding pocket.

The constraint that we impose on pocket candidates is that that they are accessible to possible ligands, which means they lie on the surface of the network. Another constraint is that they actually represent a *pocket*, that is the lines connecting them do not cut through the volume of the network. Both these constraints rely on a definition of a surface of the network. Defining the surface of a network (i. e. the concave hull of a cloud of points in 3 dimensions) is neither mathematically nor intuitively well defined. Many possible different shapes would envelop the respective structures. The method we chose to determine a unique surface is to grow a sphere around each of the beads such that a coherent solid body emerges. This *fused spheres* object[6] defines a possible surface of the network that actually is quite similar to the definition of the SAS area for proteins. We declare beads that belong to spheres that participate in this outer shared surface as *surface beads* and the remaining ones as interior beads. The method is visualized in Fig. 4.4.

Pocket pairs and triplets are now selected from the set of surface beads. The condition mentioned above that binding pockets lie on the outer surface of the network and rep-

[6]The numerical implementation was kindly made available by Lukas Engelke [Engelke, 2021]

resent pockets, rules out implausible ligand effects. It is enforced as follows. Out of all possible combinations of surface bead pairs and triplets we only consider those whose direct connecting lines do not cut any of the interior beads. In addition, there are two more constraints reducing the overall number of bead pairs to scan. We omit pairs if they are direct neighbors, or if their distance is above a certain threshold value.

4.3 Correlations at equilibrium

From MD simulations it is well known that the local motions within orthosteric and allosteric sites are highly correlated [Ma et al., 2016, VanWart et al., 2012]. Especially for allosteric networks it might thus be instructive to evaluate the correlation function between the state of the target pocket d_t and the source pocket d_s.

In the following we derive analytic expressions for the single-point and two-point probability densities, which give access to the correlations between distances in equilibrium.

We recall the harmonic approximation of the energy function in Eq. (2.3),

$$U(\delta\boldsymbol{r}) = \frac{1}{2}\delta\boldsymbol{r}^{\mathsf{T}}\underline{\mathrm{H}}\delta\boldsymbol{r}. \tag{4.1}$$

This formulation of the energy allows to describe the dynamics in terms of small deviations from equilibrium positions, $\delta\boldsymbol{r} = \boldsymbol{r} - \boldsymbol{r}^{(0)}$ and is thus equivalent to a $3N$ dimensional Ornstein-Uhlenbeck Process (OUP). Switching to normal coordinates allowed for a simpler formulation of the energy (Eq. (2.10))

$$U(\boldsymbol{v}) = \sum_{\nu} \lambda_\nu v_\nu^2. \tag{4.2}$$

The partial differential equation that describes the time evolution of the probability density function of the positions' deviations is the overdamped version of the *Fokker-Planck equation*, also known as the *Smoluchowski equation* [Gardiner et al., 1985] and reads

$$\frac{\partial P}{\partial t} = \sum_{k=1}^{3N} \frac{\partial^2 P}{\partial v_\nu^2} + \frac{\partial}{\partial v_\nu} 2\lambda_\nu^2 v_\nu P. \tag{4.3}$$

The solution in equilibrium is given by a multivariate *Gaussian distribution*:

$$P_{\text{eq}}(\boldsymbol{v}) = \prod_{\nu=1}^{3N} \left[\left(\frac{\lambda_\nu^2}{\pi} \right)^{1/2} e^{-\lambda_\nu^2 \boldsymbol{v}_\nu^2} \right] \tag{4.4}$$

This allows us to write the probability density function of the fluctuations of distances in the network as the single point density, and the expectation value of a certain distance $\boldsymbol{d}$ is given by

$$P(\,\boldsymbol{d}\,) = \lim_{t\to\infty} \langle \delta(\delta \boldsymbol{r}(t) - \boldsymbol{d}) \rangle_{\text{eq}}, \tag{4.5}$$

where $\delta(\cdot)$ is the *Dirac delta*. We switch now to three dimensional vectors, such that $\vec{d}_{ij} = \vec{r}_j - \vec{r}_i = \sum_{\nu=1}^{3N} (\underline{\mathrm{V}}_{j\nu} - \underline{\mathrm{V}}_{i\nu}) v_\nu = \underline{\mathrm{V}}_\nu^{ij} v_\nu$, i and j each denoting the three corresponding rows of the matrix $\underline{\mathrm{V}}$. We further introduce the abbreviation

$$\gamma = \sum_{\nu=1}^{3N} \frac{\underline{\mathrm{V}}_\nu^2}{2\lambda_\nu^2} \,. \tag{4.6}$$

$\vec{d}_0$ is the distance in the reference structure. Then the probability density for a specific distance vector is

$$\begin{aligned} P(\,\vec{d}\,) &= \int d\boldsymbol{v}\, P_{\text{eq}}(\boldsymbol{v})\; \delta \left(\sum_{\nu=1}^{N} \underline{\mathrm{V}}_\nu^{ij} v_\nu \;+\; \vec{r}_{ij} - \vec{d} \right) \\ &= \frac{1}{(2\pi)^3} \left(\frac{\pi}{\gamma} \right)^{3/2} e^{-\frac{(\vec{d}-\vec{d}_0)^2}{4\gamma}}, \end{aligned} \tag{4.7}$$

where we have used a forward and inverse Fourier transformation to evaluate the integral. If we are only interested in the distance $d = |\vec{d}|$ irrespective of the orientation of the vector, we have to integrate over all possible orientations. Choosing without loss of generality the frame of reference along the z-axis, yields:

$$\begin{aligned} P(d) &= \int_0^{2\pi} d\phi \int_0^{\infty} d\rho \int_{-1}^{1} d\cos(\theta) P(\vec{d}) \delta(|\vec{d}| - \rho) \\ &= \frac{1}{\sqrt{\pi\gamma}} \frac{d}{d_0} \exp\left[-\frac{d^2 + d_0^2}{4\gamma} \right] \sinh\left[\frac{d d_0}{2\gamma} \right]. \end{aligned} \tag{4.8}$$

The evaluation of the two point density, that is the joint probability density of two distances, is slightly more involved. We introduce the following abbreviations

$$A = \sum_{\nu=1}^{3N} \left(\frac{\mathrm{V}_{i\nu}}{2\lambda_\nu} \right)^2, \quad B = \sum_{\nu=1}^{3N} \left(\frac{\mathrm{V}_{j\nu}}{2\lambda_\nu} \right)^2, \quad C = \sum_{\nu=1}^{3N} \frac{\mathrm{V}_{i\nu}\mathrm{V}_{j\nu}}{2\lambda_\nu^2}, \quad D = 4(AB - C^2). \tag{4.9}$$

Again we have for the distance vectors $\vec{d_i}$ and $\vec{d_j}$:

$$P(\vec{d_i}, \vec{d_j}) = \left(\frac{1}{2\pi} \right)^3 \left(\frac{1}{D} \right)^{\frac{3}{2}} \exp\left[-\frac{1}{D} \left(A\vec{d_j}^2 + B\vec{d_i}^2 + C\vec{d_j} \cdot \vec{d_i} \right) \right]. \tag{4.10}$$

The marginalization over the possible orientations of both vectors gives the probability density for $d_i = |\vec{d_i}|$ and $d_j = |\vec{d_j}|$, which reads

$$P(d_i, d_j) = \frac{1}{\pi} \left(\frac{1}{D} \right)^{\frac{1}{2}} \frac{d_i d_j}{C} \exp\left[-\frac{1}{D} \left(A d_j^2 + B d_i^2 \right) \right] \sinh\left[\frac{C}{D} d_j d_i \right]. \tag{4.11}$$

We can use the joint density and the marginals to calculate the correlation function [Papoulis and Pillai, 2002] of a pair of distances in the network

$$P(d_\mathrm{s}, d_\mathrm{t}) - P(d_\mathrm{s}) P(d_\mathrm{t}) \tag{4.12}$$

To test whether this observable is a good descriptor for allosteric long-range interactions we determined the correlation function for an untrained (random) network and two trained derivatives of this network, trained for a symmetric and an antisymmetric effect from the source to the target pocket. We used the networks that are readily available from [Flechsig, 2017]. The data is shown in Fig. 4.5.

We can directly observe that there are large correlations between the source and the target distance in all networks. However, there is no qualitative difference between the trained and the untrained network. Therefore, we can conclude that the equilibrium correlations between the fluctuations of the source and target distance are not a descriptor that should be used to quantify allostery. We therefore decided to take one step back and first solve for the minimum-energy path before coupling the system to a heat bath.

A comparison of the impact of the fluctuations caused by the heat bath and the actual allosteric changes in distance is given in Appendix A.7.

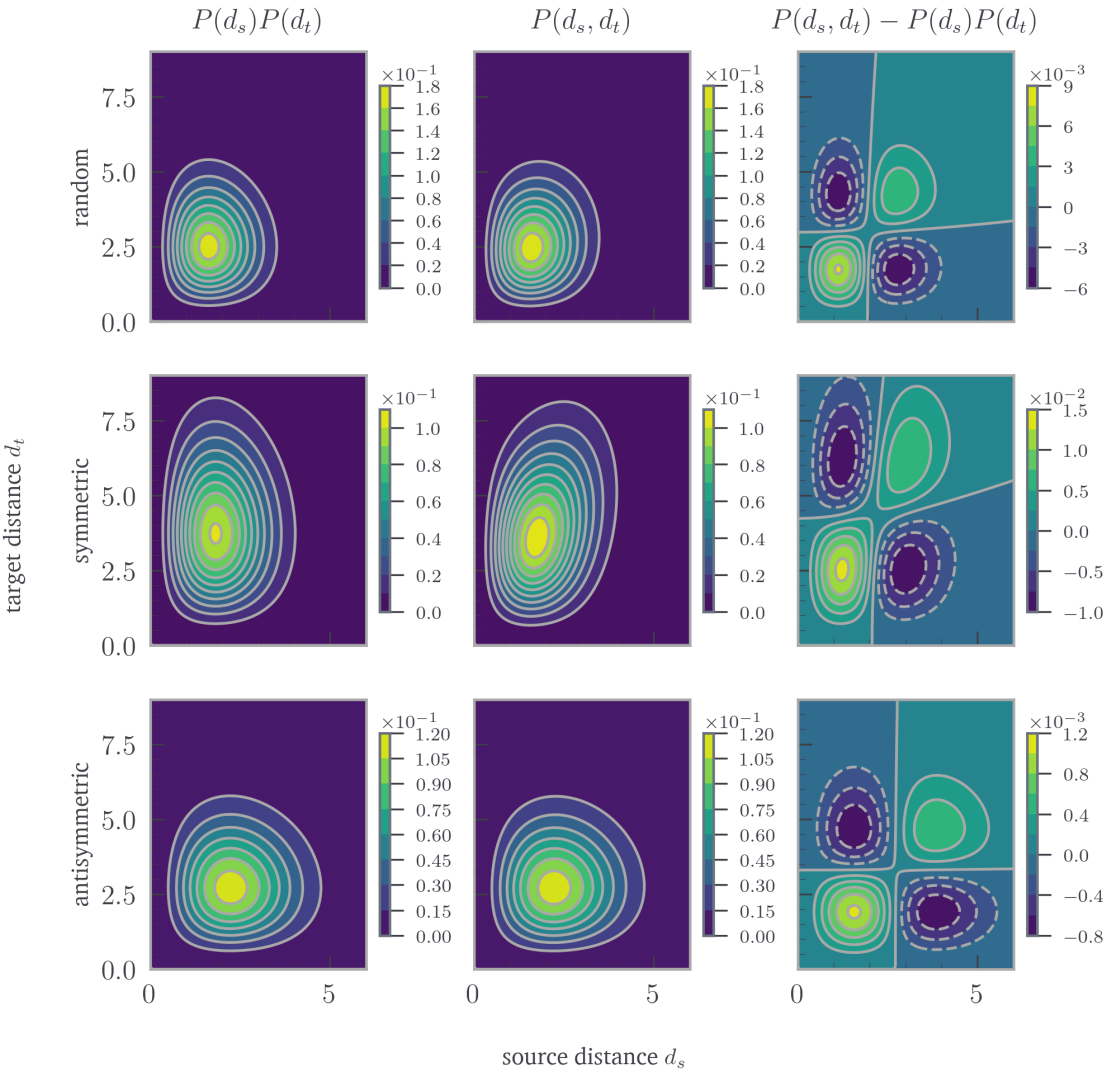


Figure 4.5: Joint and individual probability densities as well as the correlation function of the distances between source and target pocket beads for the three different networks in [Flechsig, 2017]. No qualitative difference between the trained and untrained networks is observable.

Chapter 5

Allosteric Response

There are many different ways to define and determine the response of an ENM to an external stimulus. A first order approximation to calculate the effect of binding is Linear Response Theory (LRT), as described in Section 2.5.1.

The probably most intuitive method, however, is to simply integrate Newton's equations of motion with an additional external force mimicking the effect of a ligand binding event [Flechsig, 2017]. In the overdamped limit with friction γ and spring κ constants set to unity[1] the equations of motion read

$$\begin{aligned} \dot{\vec{r}}_i &= -\nabla_{\vec{r}} U + \vec{f} \\ &= -\sum_{j \neq i} \underline{\mathrm{A}}_{ij} \left(1 - \frac{r_{ij}^{(0)}}{|\vec{r}_i - \vec{r}_j|}\right) (\vec{r}_i - \vec{r}_j) + \vec{f}_i \end{aligned} \tag{5.1}$$

Another straightforward approach would be to apply gradient descent with the full ENM potential (Eq. (2.14)) to find equilibrium positions after a subset of them has been constrained to new positions. We can discretize time for small step sizes ϵ, and iteratively obtain new positions as follows:

$$\boldsymbol{r}_{i+1} = \boldsymbol{r}_i - \epsilon \sum_{j \neq i} \underline{\mathrm{A}}_{ij} \left(1 - \frac{r_{ij}^{(0)}}{|\vec{r}_i - \vec{r}_j|}\right) (\vec{r}_i - \vec{r}_j)\,, \tag{5.2}$$

[1] The explicit dependence on γ and κ can also be removed when switching to natural units, where the rescaled time would read $(\kappa/\gamma)t$.

noting that absolute times are irrelevant here as we are interested only in the new configurations.

The advantage of these methods is their accuracy; with a small enough step size they have been shown to accurately describe complex nonlinear behavior of proteins upon perturbation [Togashi and Mikhailov, 2007, Togashi et al., 2010, Poma et al., 2018]. However, the price one pays for this is speed.[2] We therefore combine the advantages of both approaches by applying the linear response formalism in an iterative fashion, extending the Hessian matrix to describe correctly instantaneous conformations that have left the initial minimum.

5.1 Iterative constrained quadratic optimization

We reformulate the LRT as a constrained minimization problem. Given a set of instantaneous coordinates $\boldsymbol{r}^{(0)}$ which constitute a minimum of the corresponding energy function $U(\boldsymbol{r}, \boldsymbol{r}^{(0)})$ we are interested in new optimal coordinates $\boldsymbol{r}$ after a constraint is applied to a subset of the coordinates.

For small deviations around the instantaneous coordinates, the energy function is a simple quadratic form

$$U(\boldsymbol{r}, \boldsymbol{r}^{(0)}) = \frac{1}{2}(\boldsymbol{r} - \boldsymbol{r}^{(0)})^{\mathsf{T}} \underline{\mathrm{H}} (\boldsymbol{r} - \boldsymbol{r}^{(0)}) + U^{(0)}(\boldsymbol{r}^{(0)}), \tag{5.3}$$

where $U^{(0)}(\boldsymbol{r}^{(0)})$ is the constant term in the Taylor approximation that does not vanish for extended springs. We now rewrite the quadratic optimization in such a way that it works without explicit constraints in form of Lagrange multipliers and concurrently allows for constraining multiple beads simultaneously.

We split the positions $\boldsymbol{r}$ into two subsets, the constrained fixed ones, $\boldsymbol{c}$, and the ones sought-for, $\boldsymbol{s}$, such that

$$\boldsymbol{r} = \begin{pmatrix} \boldsymbol{c} \\ \boldsymbol{s} \end{pmatrix}. \tag{5.4}$$

Similar for the instantaneous initial coordinates,

$$\boldsymbol{r}^{(0)} = \begin{pmatrix} \boldsymbol{c}^{(0)} \\ \boldsymbol{s}^{(0)} \end{pmatrix}. \tag{5.5}$$

[2] We see a strong speedup when employing a state of the art stochastic optimizer like *ADAM* [Kingma and Ba, 2014], reducing the convergence time for the response of a ∼200 bead system from hours to minutes, which is still about two orders of magnitude slower than the iterative method we propose in the next section.

The Hessian matrix can analogously be split into four blocks,

$$\underline{\mathrm{H}} = \begin{pmatrix} \underline{\mathrm{C}} & \underline{\mathrm{B}}^{\mathsf{T}} \\ \underline{\mathrm{B}} & \underline{\mathrm{S}} \end{pmatrix}. \tag{5.6}$$

The quadratic form which is to be optimized now reads

$$U(\boldsymbol{s}, \boldsymbol{c}, \boldsymbol{s}^{(0)}, \boldsymbol{c}^{(0)}) = \frac{1}{2}\boldsymbol{s}^{\mathsf{T}}\underline{\mathrm{S}}\boldsymbol{s} + \left((\boldsymbol{c}^{\mathsf{T}} - \boldsymbol{c}^{(0)\mathsf{T}})\underline{\mathrm{B}} - \boldsymbol{s}^{(0)\mathsf{T}}\underline{\mathrm{S}}\right)\boldsymbol{s}, \tag{5.7}$$

where we dropped the constant term as it is irrelevant for the optimization.

Demanding that the gradient $\nabla_{\boldsymbol{s}} U$ is zero and switching from absolute positions to displacements $\delta\boldsymbol{c} = \boldsymbol{c} - \boldsymbol{c}^{(0)}$ and $\delta\boldsymbol{s} = \boldsymbol{s} - \boldsymbol{s}^{(0)}$ for the free and fixed beads respectively, this leads to the system of linear equations

$$\delta\boldsymbol{s} = -\underline{\mathrm{S}}^{-1}\,\underline{\mathrm{B}}^{\mathsf{T}}\delta\boldsymbol{c}, \tag{5.8}$$

that is simple to solve numerically.

We note that the splitting of the matrix is equivalent to the one used in [Zheng and Brooks, 2005b] and the implicit formulation of the response induced by deforming a binding pocket in Eq. (5.8) was also derived already in [Zheng and Doniach, 2003].

In order to determine the response of the free beads to the perturbation of the fixed ones, it is henceforth straightforward to apply the above mentioned quadratic optimization in an iterative fashion. The input consists of both, the current positions and the new positions of the constrained beads. A new Hessian matrix is calculated for the current positions, taking into account the original equilibrium lengths of the springs, as described in Eqs. (5.10, 5.11, 5.12 and 5.13). The iterative protocol for obtaining new optimal coordinates then reads

$$\boldsymbol{r}(k+1) = \inf_{\boldsymbol{r}} \left[\boldsymbol{r}^{\mathsf{T}}\underline{\mathrm{H}}\boldsymbol{r} | \boldsymbol{r}_{i\in S}(k+1)\right], \tag{5.9}$$

where the Hessian matrix is evaluated at the positions of the k-th step $\boldsymbol{r}(k)$ and $\boldsymbol{r}_{i\in S}(k+1)$ corresponds to the constrained positions at the next step, $\boldsymbol{c}$.

5.2 Extended Hessian matrix

5.2.1 Extended springs

As soon as the initial minimum, where the springs are at equilibrium, is left, the approximation of the potential with the Hessian matrix described in 2.1 becomes invalid. For the derivation of the matrix super-elements in Eq. (2.16) the equivalence between the instantaneous and the equilibrium spring lengths simplifies the entries dramatically. In the following the general elements are given. We are expanding the potential energy function here around the constrained minimum positions. The diagonal super-elements read

$$H_{kl}^{\alpha\beta} = \frac{\partial^2 U}{\partial_k^\alpha \partial_l^\beta} = -\frac{r_{kl}^0}{r_{kl}^3} r_k^\alpha r_l^\beta \tag{5.10}$$

and

$$H_{kl}^{\alpha\alpha} = \frac{\partial^2 U}{\partial_k^\alpha \partial_l^\alpha} = -\left[1 + \frac{r_{kl}^0}{r_{kl}^3}\left((r_k^\alpha)^2 - r_{kl}^2\right)\right]. \tag{5.11}$$

where k, l are the indices of the beads and α, β the ones of the xyz- coordinates, respectively. The diagonal super-elements are

$$H_{kk}^{\alpha\beta} = \frac{\partial^2 U}{\partial_k^\alpha \partial_k^\beta} = \sum_{l=1}^{N} \frac{r_{kl}^0}{r_{kl}^3} r_k^\alpha r_l^\beta = -\sum_{l=1}^{N} H_{kl}^{\alpha\beta} \tag{5.12}$$

and

$$H_{kk}^{\alpha\alpha} = \frac{\partial^2 U}{\partial_k^\alpha \partial_k^\alpha} = \sum_{l=1}^{N} \left[1 + \frac{r_{kl}^0}{r_{kl}^3}\left((r_k^\alpha)^2 - r_{kl}^2\right)\right] = -\sum_{l=1}^{N} H_{kl}^{\alpha\alpha}. \tag{5.13}$$

If we set r_{kl} to $r_{kl}^{(0)}$, the known formulas are recovered.

5.2.2 Repulsive beads

During the training of the networks (described in Section 4.2.2) we occasionally observed nonphysical behavior; regions in the network that could be described by different "arms" were passing through each other. This is an artifact that can not be prevented within a representation confined only to springs as these arms are actually not connected to each other. It represents an artifact caused by the emergence of large motions during the train-

ing, which violate the excluded volume principle.[3] In order to prevent parts of the network to overlap we equip the beads with an additional repulsive potential, represented by the Weeks-Chandler-Anderson (WCA) potential that is also used in the theory and computer simulations of liquids [Chandler et al., 1983, Bishop, 1984, Deiters and Randzio, 1995].

It corresponds to the repulsive part of the Lennard-Jones (LJ) potential, shifted by the energy well depth ε and truncated at the minimum, such that σ does not describe the zero-crossing of the energy anymore. The distance between two beads i and j is $r_{ij} = |\vec{r}_{ij}|$ and the pairwise LJ energy reads

$$\phi(|\vec{r}_{ij}|) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{|\vec{r}_{ij}|} \right)^{12} - \left(\frac{\sigma}{|\vec{r}_{ij}|} \right)^{6} \right] + \varepsilon & \text{for } |\vec{r}_{ij}| \leqslant 2^{\frac{1}{6}} \sigma \\ 0 & \text{for } |\vec{r}_{ij}| > 2^{\frac{1}{6}} \sigma \end{cases}, \tag{5.14}$$

thus the full LJ energy is

$$U_{\mathrm{LJ}}(|\vec{r}_{ij}|) = \sum_{j>i}^{3N} \phi(|\vec{r}_{ij}|). \tag{5.15}$$

We will now derive the Hessian elements for an expansion of Eq. (5.15). The prefactors $\phi\prime$ and $\phi\prime\prime$ are the derivatives w.r.t the absolute of the vector r_{kl} and read

$$\phi\prime\prime(r_{kl}) = \frac{4\varepsilon}{r_{kl}^2} \left[\left(156 \frac{\sigma}{r_{kl}} \right)^{12} - \left(42 \frac{\sigma}{r_{kl}} \right)^{6} \right] \qquad \phi\prime(r_{kl}) = -\frac{4\varepsilon}{r_{kl}} \left[\left(12 \frac{\sigma}{r_{kl}} \right)^{12} - \left(6 \frac{\sigma}{r_{kl}} \right)^{6} \right]. \tag{5.16}$$

Then the off-diagonal super elements with different cartesian coordinates α, β read

$$\frac{\partial^2 U_{\mathrm{LJ}}}{\partial_k^\alpha \partial_l^\beta} = -\left[\phi\prime\prime \frac{r_{kl}^\alpha r_{ij}^\beta}{r_{kl}^2} - \frac{\phi\prime}{r_{kl}} \frac{r_{kl}^\alpha r_{ij}^\beta}{r_{kl}^2} \right] \tag{5.17}$$

and for the diagonal elements in the off-diagonal super elements it follows that

$$\frac{\partial^2 U_{\mathrm{LJ}}}{\partial_k^\alpha \partial_l^\alpha} = -\left[\phi\prime\prime \left(\frac{r_{kl}^\alpha}{r_{kl}^2} \right)^2 - \frac{\phi\prime}{r_{kl}} \left(\left(\frac{r_{kl}^\alpha}{r_{kl}^2} \right)^2 - 1 \right) \right]. \tag{5.18}$$

The diagonal super-elements obey the same symmetry as Eq. (2.17).

[3] This artifact does not occurr for protein-derived networks.

We include the repulsive potential in the Hessian matrix $\underline{\mathrm{H}}$ whenever beads that are not connected by a spring come closer than $r_{\mathrm{rep}} = 2^{\frac{1}{6}}\sigma$. The values of r_{rep} and ε are given in the overview Table A.2

Notably, in the evaluation of the response of proteins these additional terms are *not* required, we only use them for self-trained networks.

5.3 Removing rigid body motions

The Hessian matrix is singular because it has a nullspace $\mathcal{N}(\underline{\mathrm{H}})$ which is spanned by the vectors describing infinitesimal rigid body motions. Depending on the number of constraints applied to the system the solution to Eq. (5.8) is not always unique, i. e. the system is underdetermined. A simple algebraic trick solves this problem.

Consider the square matrix $\underline{\mathrm{M}} \in \mathbb{R}^{m\times m}$ which does not have full rank, its nullity is n. Its nullspace is $\mathcal{N}(\underline{\mathrm{M}}) \in \mathbb{R}^{m\times n}$ and its row space (or range) is $R(\underline{\mathrm{M}}^{\mathsf{T}}) \in \underline{\mathrm{R}}^{m\times(m-d)}$. All vectors in $\mathcal{N}(\underline{\mathrm{M}})$ are orthogonal to all vectors in $\mathcal{R}(\underline{\mathrm{M}})$. If we add the matrix given by the outer product of a basis vector from the nullspace to $\underline{\mathrm{M}}$, i. e.

$$\tilde{\underline{\mathrm{M}}} = \underline{\mathrm{M}} + \boldsymbol{x} \otimes \boldsymbol{x}^{\mathsf{T}} \tag{5.19}$$

we increase the rank by one. Applying this for each of the basis vectors of $\mathcal{N}(\underline{\mathrm{M}})$[4] gives $\underline{\mathrm{M}}$ full rank [Blázquez et al., 1996]. This however, will not change any result where the matrix is applied to a vector from the row space, since

$$\begin{aligned} \tilde{\underline{\mathrm{M}}}\boldsymbol{y} &= \underline{\mathrm{M}}\boldsymbol{y} + (\boldsymbol{x} \otimes \boldsymbol{x}^{\mathsf{T}})\boldsymbol{y} \\ &= \underline{\mathrm{M}}\boldsymbol{y} + (\underbrace{\boldsymbol{x} \cdot \boldsymbol{y}}_{=0})\boldsymbol{x} \qquad\qquad \forall\, \boldsymbol{x} \in \mathcal{N}(\underline{\mathrm{M}}) \\ &= \underline{\mathrm{M}}\boldsymbol{y}. \end{aligned} \tag{5.20}$$

[4]Note that a standard eigendecomposition routine is not ideal for a fast calculation of the eigenvectors that correspond to the six zero eigenvalues of the Hessian matrix. Using a shift and invert spectral transformation from *ARPACK* [Yang et al., 1997], where we choose the number of eigenvalues we are interested in and give a starting point for the search, is considerably faster. The corresponding implementation in *scipy* [Virtanen et al., 2020] does not always converge for eigenvalues that are of the same magnitude as the numerical precision (zero eigenvalues). We overcome this limitation by adding a small noise to the matrix whose nullspace is to be determined.

In our case, we do not apply this rank extension to the Hessian matrix, but instead only to the sub-block containing the free positions $\boldsymbol{s}$ in Eq. (5.8). Constraining the positions of beads in $\boldsymbol{c}$ would otherwise conflict with the inhibited degrees of freedom.

The resulting matrix has full rank rendering the linear system (Eq. (5.8)) uniquely solvable. The solution by construction does not contain any rigid body motions.

There is also a physical interpretation to the nullspace extension of the Hessian matrix. The way the basis vectors of the nullspace are combined to span the additional matrix is similar to how the dyades of eigenvectors (scaled with the eigenvalues) are summed for creating the full matrix. Using the interpretation of the eigendecomposition of a Hessian matrix, where the collective motions happen along high dimensional parabolae, we can argue that we are confining the full system by parabolae.

It is indeed even possible to identify the basis vectors of the nullspace explicitly. In the Cartesian basis with 3 dimensions the vectors generating the translation of a body with N beads can be written conveniently as $3N$ dimensional vectors:

$$\boldsymbol{\tau}(x) = \begin{pmatrix} \vec{e}_x \\ \vdots \\ \vec{e}_x \end{pmatrix}, \quad \boldsymbol{\tau}(y) = \begin{pmatrix} \vec{e}_y \\ \vdots \\ \vec{e}_y \end{pmatrix}, \quad \boldsymbol{\tau}(z) = \begin{pmatrix} \vec{e}_z \\ \vdots \\ \vec{e}_z \end{pmatrix}, \tag{5.21}$$

with the three Cartesian basis vectors

$$\vec{e}_x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \vec{e}_y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \vec{e}_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{5.22}$$

Given the positions of the beads

$$\boldsymbol{r} = \begin{pmatrix} \vec{r}_1 \\ \vdots \\ \vec{r}_N \end{pmatrix} \tag{5.23}$$

the basis vectors generating the (infinitesimal) rotation of this body can be constructed as follows

$$\boldsymbol{\rho}(\alpha) = \begin{pmatrix} \vec{e}_\alpha \times \vec{r}_1 \\ \vdots \\ \vec{e}_\alpha \times \vec{r}_N \end{pmatrix}, \tag{5.24}$$

where α indicates $\{x, y, z\}$.

Both methods are implemented and show similar results. Notably, the second method is faster. If we constrain more than two beads (essentially always true for the binding sites in proteins) the removal of rigid body motions can be omitted completely, speeding up the algorithm tremendously.

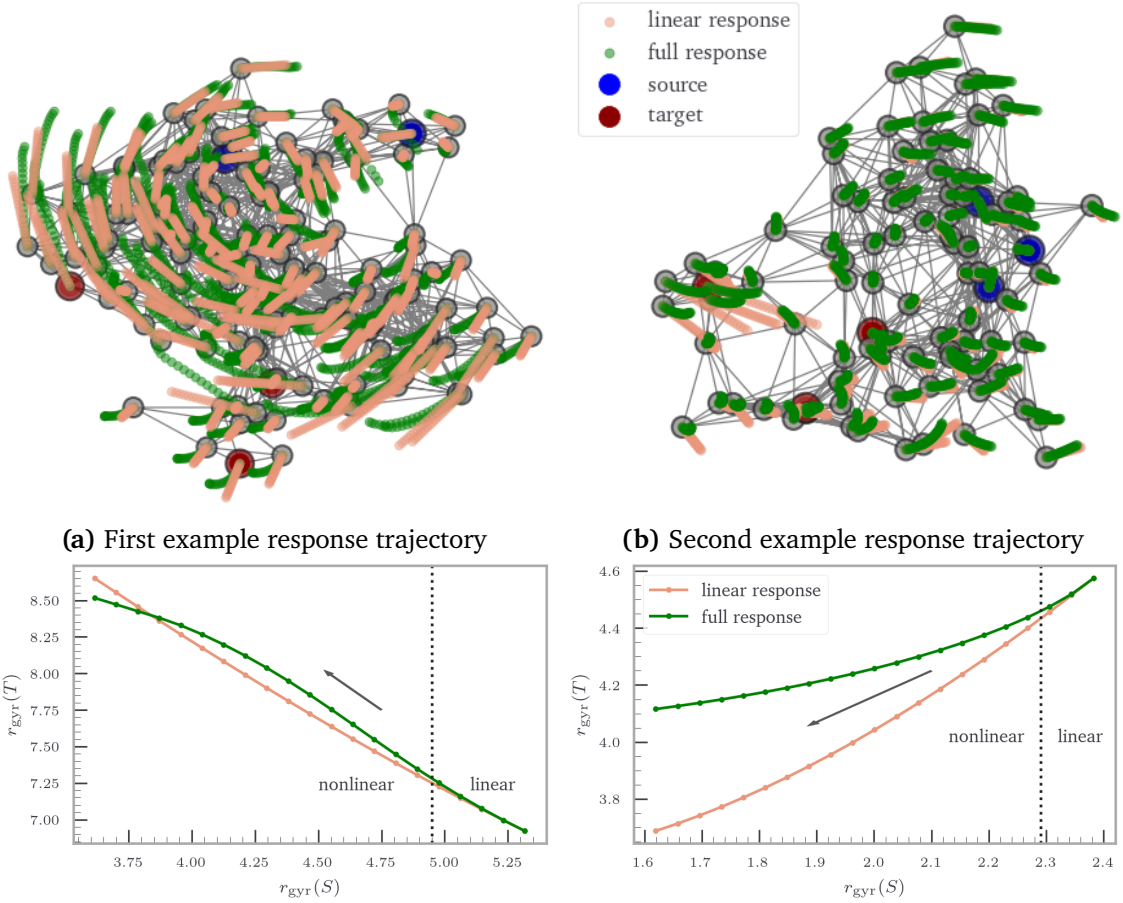
5.4 Nonlinearities

We discussed in Section 2.5.2 that the ENM is inherently nonlinear. Nevertheless, it is a model that can only be valid for describing the conformations of a protein maintaining its instantaneous fold. It is within these boundaries where the ENM, mainly used in hybrid MD simulations in this regard, has been shown to describe complex nonlinear conformational motions which can be responsible for the allosteric long-range effects [Togashi et al., 2010].

Using the iterative linear response algorithm developed in Section 5.1 we now try to validate that the constrained minimum energy path indeed reproduces such nonlinear behavior, and thus represents a fast and accurate alternative for systems where a single-step LRT fails.

The perhaps simplest way to visualize nonlinear behavior is do depict the difference between the full and the linear solution to the same problem. We compute the response for the same sequence of input perturbations by updating the Hessian matrix in each step and using the initial Hessian matrix in all steps. With the response for the first step being equivalent, we see a growing deviation for larger perturbations. Two examples are shown in Fig. 5.1a and Fig. 5.1b. Especially for rotational motions (Fig. 5.1a) a description with a single vector cannot be complete. Notice how the lower arm of the network, constituting part of the target pocket, deviates from the linear response already for small inputs. Similarly, the main difference between the linear and full response lies, (see Fig. 5.1b), in the motion of beads exerting the allosteric response at the target pocket, where a precise rearrangement is actually required the most [Daily and Gray, 2007] This reinforces the demand for a description that accurately resolves nonlinear motion. This non-uniformity in the capability of linear response to describe the allosteric motion bears the risk to overinterpret the results obtained with the linear method—a high overall accuracy might hide the strong local deviations that occur at the target site, which is apparently extremely susceptible to deviations.

The differences between the linear and the full response are also evident in the observables we used for training the networks, i. e. the radii of gyration r_{gyr} of the source and target site. Figs. 5.1c and 5.1d show these radii for two examples, illustrating especially that the



(a) First example response trajectory

(b) Second example response trajectory

(c) Radius of gyration (first example)

(d) Radius of gyration (second example)

Figure 5.1: Examples where linear response fails to describe conformational motions. While identical in the first step, for multiple steps the deviations can grow substantially. Especially small local motions at the target pocket are not accurately described by the linear response. Green lines and circles depict the full response, the orange ones correspond to the linear response. Note that the radius of gyration is a nonlinear observable itself, thus both, the lines obtained via the linear and the nonlinear response, show non straight lines.

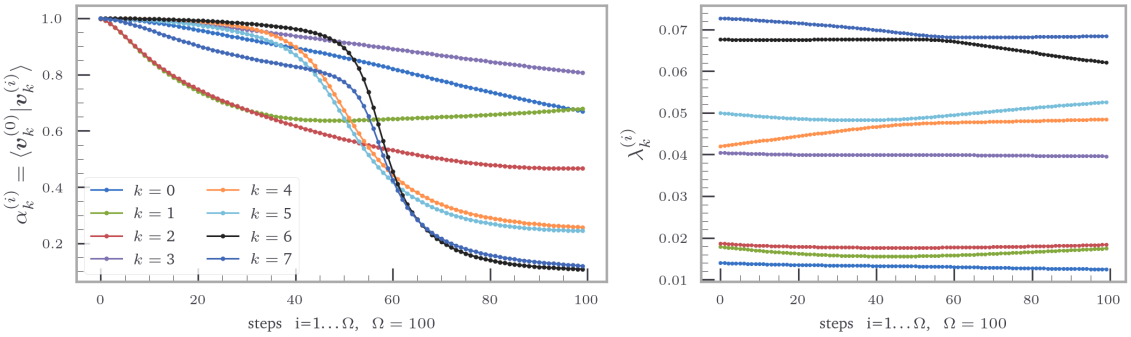


Figure 5.2: Manifestations of nonlinearity in the response. The consecutive scalar product of eigenvectors, (see Eq. (5.25)), shows that these rotate during the response, indicating that the direction of the flattest path in the energy landscape changes. Eigenvalues change their value and can even swap, which means the steepness of the potential changes. These two observables tell us that we are leaving a valley with different slopes and orientations during the allosteric response.

points where the linear response diverges from the full response can differ substantially between the networks.

As we have access to the full sequence of Hessian matrices we can explain the observed nonlinearities using spectral insights along the response path. The local motion inside the first potential energy minimum is described by a combination of the curvature and the orientation of the high dimensional instantaneous parabolae. The change of these two observables is traced in Fig. 5.2, for the first six non-zero eigenvalues. The eigenvalues in the right panel of Fig. 5.2 exhibit non-monotonic changes, indicating that the slope of the parabolae against which the system is dragged varies substantially along the lowest-energy path.

The changes of the eigenmodes are most easily analyzed in terms of the rotation these undergo during the rearrangements of the network. We therefore plot the scalar product of the first six non-zero eigenvectors, for their i-th step against the initial one,

$$\alpha_k^{(i)} = \langle \boldsymbol{v}_k^{(0)} | \boldsymbol{v}_k^{(i)} \rangle, \tag{5.25}$$

see the left panel of Fig. 5.2. The example depicted here shows the different qualitative changes that may occur during the response, them being almost zero for modes 3 and 4, a steady rotation as mode 0 does, or strong but abrupt rotations like modes 1, 2 and 5. In Figs. 5.3 and 5.4 we provide the statistics of these two observables, the change of the eigenvalues and the rotation of the eigenmodes during the response, respectively. Apparently, the effect of nonlinearity can differ strongly in magnitude between the studied networks.

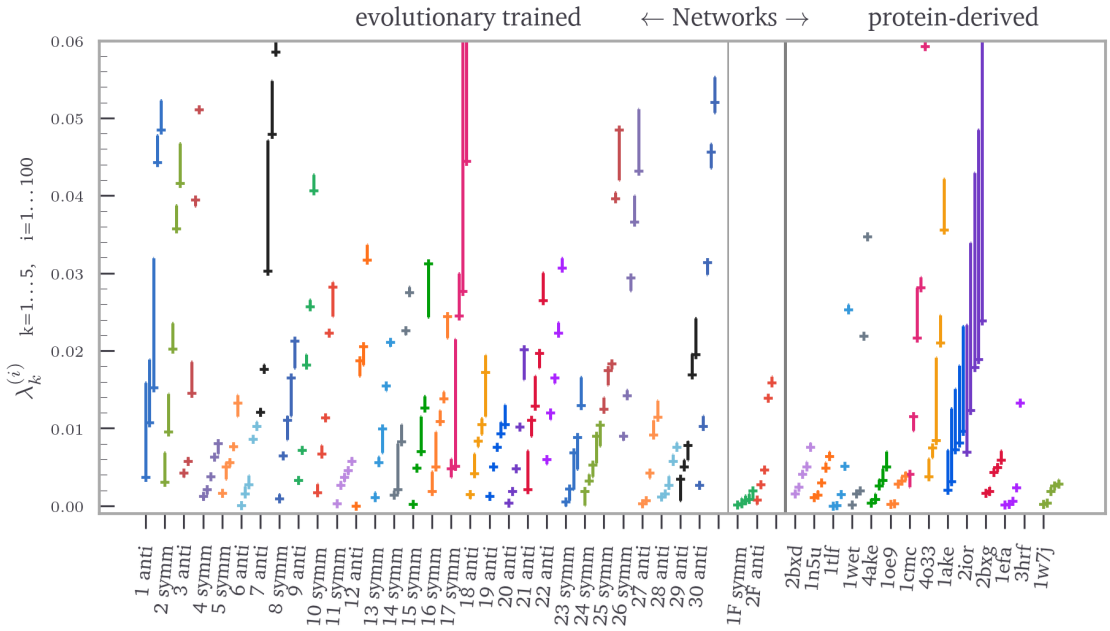


Figure 5.3: Statistics over the first five non-zero eigenvalue variations during the response for all artificial (left) and protein-derived (right) networks. The initial values are denoted with crosses, the range they cover during the response is plotted.

The interpretation that accompanies these changes is simple but informative: Although we do not observe any zero-crossing of eigenvalues, meaning that we remain within the initial energy basin, the complex shape and anharmonicity of this basin carries enough high-dimensional information to encode complex motions and allosteric behavior. Considering that proteins actually perform motions that are thought to range across multiple minima, this might appear as surprising. However, it is conceivable that we are describing just the first part of the allosteric transition, i. e. the path of leaving the initial basin. The complete path, described by the full potential (see Eq. A.1), may thereafter potentially lead to a new basin. For allostery, this lowest-energy exit-path might already be of great importance.

We can thus conclude that that for the existence of nonlinear motion in ENM no barrier crossing is required, i. e. it can also occur within the first energy basin. This brings us one step closer towards a fast and accurate method for analyzing the motion of proteins, which are inherently nonlinear even without considering partial or complete un- or refolding.

No major barrier crossing during macroscopic responses is also highly consistent with the the observed sparsity of metastable states during the relaxation simulations with the full ENM potential in [Togashi et al., 2010]. Almost all of the randomly perturbed configurations found their way back to the initial minimum, very few got stuck in metastable states.

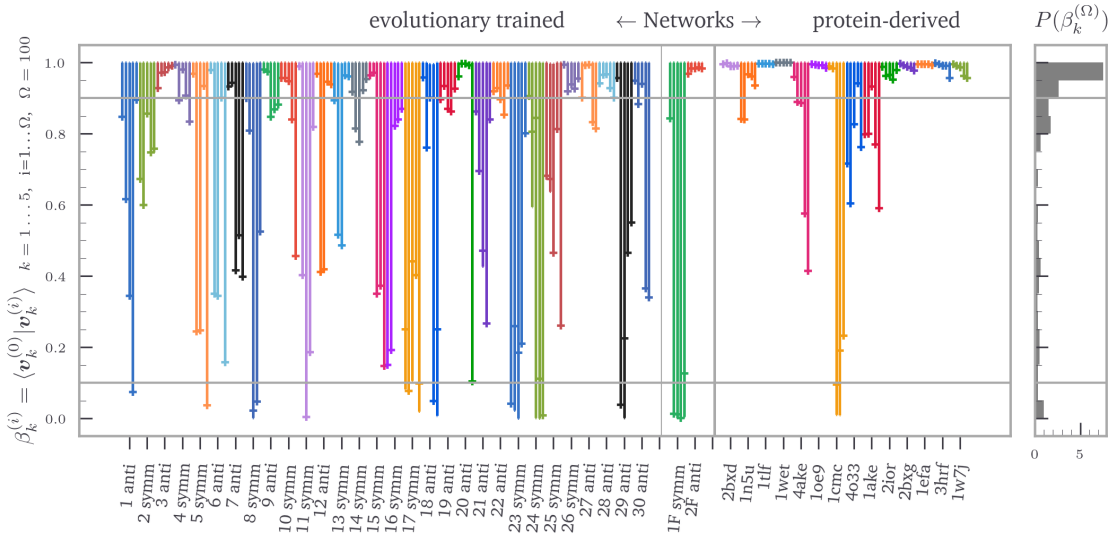


Figure 5.4: Statistics over the rotation of the first five non-zero eigenmodes during the response for all artificial (left) and protein-derived (right) networks. The initial values are denoted with crosses, the range they cover during the response is plotted. It seems that there are three classes of networks, (i) those that do not exhibit eigenvector rotation, for which a description using only the first minimum is sufficient, (ii) those with an intermediately strong rotation, and (iii) those networks that experience rotation until the modes are essentially orthogonal to the initial ones. These classes are indicated with grey lines.

The elastic energy corresponding to the trajectories was found to deviate from a quadratic dependence for the nonlinear path, see Appendix A.2.

Our method therefore enters between the naïve linear response and a full-fledged MD simulation, and is apparently detailed enough to describe complex motions but at the same time fast enough to allow for complete scanning, i. e. computing responses of all possible combinations of binding pockets—a task which we will address in Section 5.6.

We briefly recall and clarify the findings about nonlinearities in motions of proteins discussed in Section 2.5.2 and assess the results in light of these. Comparing against the empirical conformational changes obtained via PCA of available crystal structures it was shown that:

[Zheng and Doniach, 2003] Linear response successfully reproduces conformational changes upon deformation of the binding pocket, and a single soft mode adequately describes the transition well of F1-ATPase and myosin. However, kinesin shows no single-mode overlap with the response and a (single-step) linear response cannot reproduce the conformational motion.

[Zheng and Brooks, 2005b] In both myosin and kinesin show a strong coupling between a pocket mode (obtained by analyzing the subsystem of the pocket in isolation) and the global modes, although in the case of kinesin more modes are required. In the case of kinesin not the softest, but actually a slightly stiffer mode (i. e. #10) shows the highest overlap with both, the conformational change and the highest coupling to the local pocket mode.

[Togashi and Mikhailov, 2007] In both, F1-ATPase and myosin nonlinearities are present in the relaxation paths. Nevertheless, a single soft-mode description captures the qualitative features.

[Togashi et al., 2010] In kinesin the nonlinearities in the relaxation are so strong that normal mode analysis is not possible within the first energy basin no matter how many soft modes are combined.

Many plausible explanations for the origin of such nonlinearities have been anticipated so far:

[Ravasio et al., 2019] A bending in phase space caused by the motion following multiple modes simultaneously.

[Togashi et al., 2010, Kim et al., 2019] Branching of the potential energy landscape that leads to motions differing between the branches.

[Miyashita et al., 2003] Combinations of modes (as above) or local unfolding.

[Frauenfelder et al., 1991, Daily and Gray, 2007] The existence of multiple minima in the underlying energy landscape.

We can reproduce nonlinearities similar to those observed above and add another interpretation that unifies some of those mentioned above. Thinking along the spectral properties of the Hessian matrix underlying the changing network it is conceivable that the normal modes rotate during the response. This interpretation is independent of the number of modes required to show a bending in phase space, even for the case that a single mode is dominating the allosteric response initially, it can participate in the rotation that modes undergo during the response and lead to nonlinearities sooner or later. As we do not observe negative, (not even zero), eigenvalues and do not accommodate for changes in the connectivity of the ENM, the latter two cases are out of scope for our method.

5.5 Non-reciprocity

As we argued for the hypothesis in Chapter 3, in order to provide a switch-like control, allostery should to work in a non-reciprocal manner. This means that the information propagation in form of the conformational change should be transported unidirectionally from the active site to the allosteric site, and should not travel equally in the reverse direction. We see that in fact that this is exactly what occurs in the response of the networks, there is a clear qualitative and quantitative difference between the response to a perturbation at the source and the target site. An example is shown in Fig. 5.5, where again we chose as observable the radii of gyration of the two binding sites. The ratio of these two radii differs greatly between the standard and the inverse pulling setting,[5] Fig. 5.6 shows a histogram over these ratios for all studied networks and demonstrates that non-reciprocity is indeed the norm rather than an exception. For an exemplary full trajectory of a protein we point to the Appendix, Fig. A.3.

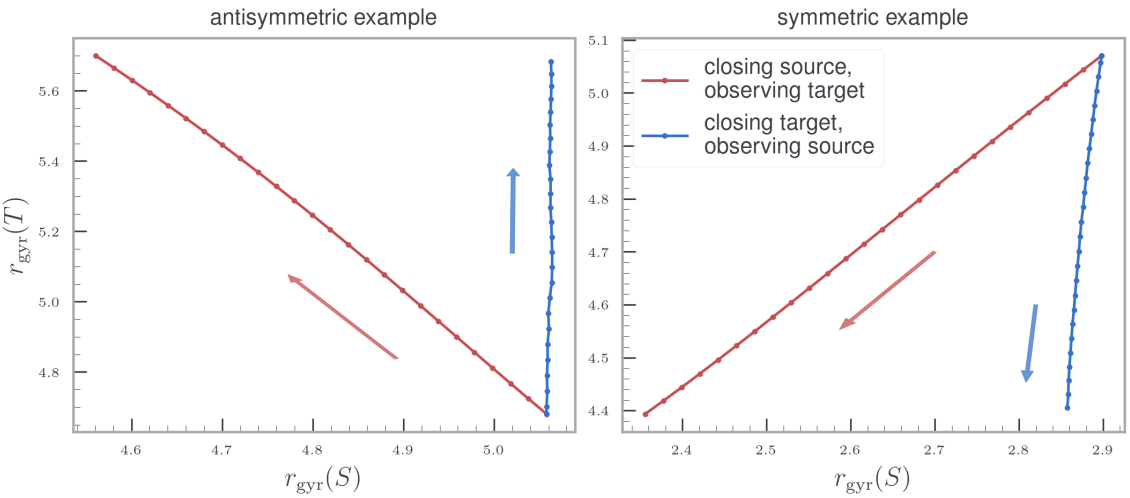


Figure 5.5: The curves show the radii of gyration of source and target pocket during the response, for pulling on the source beads (red curve) and pulling on the target beads (blue curve), examples for both an symmetric (left) and an antisymmetric (right) network are shown. The inverse pulling induces essentially zero response at the source, indicating that the allosteric effect in this network in non-reciprocal. This example is representative for essentially all our trained networks. A trajectory displaying the non-reciprocity of a protein is given in the Appendix, see Fig. A.3.

Thus, inverse pulling on the target beads does not have the same reciprocal effect on the source beads as compared to the original setting. We provide a simple explanation for how

[5] For the input of the inverse setting we use exactly the positions we obtained during the response of the standard setting.

this can be encoded in the energy landscape of the ENM and how we resolve it with the constrained optimization response, Eq. (5.8). The constraint on the fixed beads and the subsequent energy minimization we enforce on the source beads defines a hypersurface on which the positions of the free beads must lie. The free beads move along the lowest energy basin that is available on this hypersurface. Suppose we now, on the contrary, constrain the target beads to follow precisely their former response. In that case, the hypersurface will generally be different, and so will be the minimum path because the constraint is different, essentially orthogonal to the former one. Thus, the trajectories of the respective free beads will generally not be the same. This follows also directly from Eq. (5.8): Rearranging the equation for a different set of fixed and free beads, we obtain entirely different blocks of matrices required for the determination of the free beads.

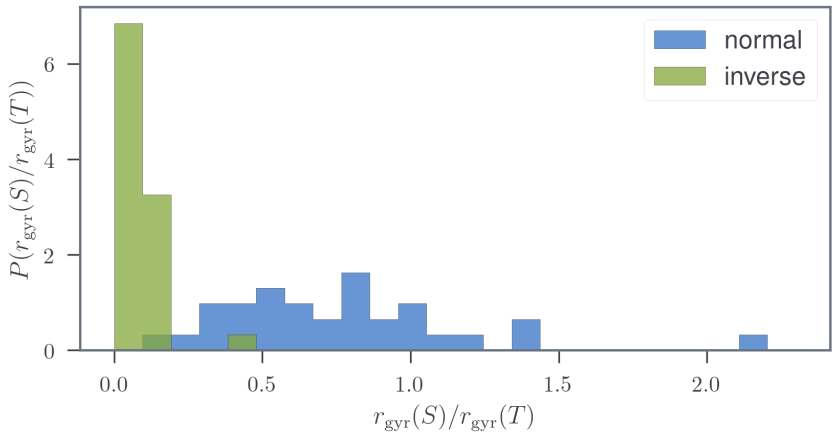


Figure 5.6: Statistics over the ratios of the final radii of gyration of the two binding sites, for the standard setting in blue and the inverse setting in green. Histograms are taken over all studied networks, and demonstrate that here, non-reciprocity is the norm rather than an exception.

For remotely regulated networks, one might want to distinguish between non-reciprocal switches and networks that exhibit a symmetric, cooperative effect. For the latter, e. g. observed in the symmetric tetramer hemoglobin, the mechanism appears to be reciprocal at first. Reciprocity is intuitively expected whenever there is a single soft mode, reminiscent of an Inextensional Mechanism (IM), underlying the allosteric motion. The mechanism might still correspond to a high-dimensional trajectory, but it is effectively a one-dimensional motion as a single parameter suffices to describe it. For such a motion, it does *not* matter from which side it is perturbed; there is only a single degree of freedom which the response can follow. As long as the perturbation happens along the mode it will be reciprocal. This argument is in line with the findings of Yan et al., who report solely single soft-mode mechanisms for networks which they trained for cooperativity and where *both active and allosteric*

sites play a symmetric role [Yan et al., 2018]. Returning to hemoglobin, where the binding sites are identical, it does not matter which is occupied first; either one induces the change that allows the next one to bind O_2 more easily. Notably, a global soft mode (#2) has been reported in hemoglobin that shows a high degree of symmetry and causes identical displacements at all four binding sites [Perahia and Mouawad, 1995]. However, there is a specific set of residues in the vicinity of the binding sites, that is shown *not* to participate in the global soft mode but reacts sensitively to the binding of O_2 and thus plays a major role in inducing the allosteric response [Xu et al., 2003]. Therefore we conclude that, although the existence of a global soft mode is essential here, it definitely cannot explain the entire picture.

Unidirectional allostery is known to exist in proteins, shown by both MD simulations [Guo and Zhou, 2016] and in experiments [Hosokawa et al., 2021]. Models employing unidirectional allostery explain and reproduce experimental data better and with fewer additional assumptions than the bidirectional counterpart [Koda and Saito, 2020]. This effect is also well-known in the field of metamaterials [Coulais et al., 2017].

5.6 Specific response scanning

The considerable gain in speed that our iterative quadratic optimization formulation entails for computing the full response allows for large-scale screening applications. The (iterative) full response provides a substantial advantage compared to single-step linear response, as for proteins the precise rearrangement at the active site it thought to be important [Daily and Gray, 2007].

We assume now that the target site is known. The aim is to allosterically control and rearrange this target in a specific manner. The allosteric (source) pocket is in turn unknown and to be determined.

Note that enforcing the wanted response at the target site and observing the rest of the network to deduce the allosteric site is not possible for the trained networks as well as the protein-derived networks, because the motion is non-reciprocal, see Fig. 5.6. Although it might work in cases where the mechanism is indeed reciprocal, pinning down an allosteric site by assuming a bidirectional coupling between them can thus *not* be in general applicable [Tee et al., 2018, Ni et al., 2021b]. The method we propose works irrespective of reciprocity.

In the following we will demonstrate the feasibility of a pre-scan to be used in drug design. For this purpose, we define the configuration that is targeted by drug design as the final

configuration at the target site, which we also achieve by perturbing the (in our case known) source pocket. Then, we scan all possible pocket pairs and triplets, for proteins also the large pockets with a 100 step iterative linear response. Nearest neighbors of the target are omitted from the scan. The computation time for all 14 proteins and 33 trained networks was a few hours on a medium sized cluster, negligibly short compared to MD-based scans requiring weeks to months of computation on the same hardware. The exact number of scanned pockets per protein is given in Table A.2 and A.1.

We now define the initial configuration of the target site ($A = T$) as $\mathbf{T}^0$. The final configuration at the end of the response are $\mathbf{T}_i^{\text{fin}}$ and $\mathbf{T}_*^{\text{fin}}$, corresponding to perturbing the i-th candidate pocket and the true source pocket, respectively. Only the actual configuration at the target site matters, the response of the rest of the network is irrelevant for drug design applications. We therefore apply an optimal rototranslation [Kabsch, 1976] in order to optimally overlay the respective target configurations.

In drug design applications $\mathbf{T}_*^{\text{fin}}$ is the desired response. We introduce the relative magnitude of the response to closing the pair/pocket i as

$$0 \leq \Delta_i \equiv \frac{\left\|\mathbf{T}_i^{\text{fin}} - \mathbf{T}^0\right\|_2}{\left\|\mathbf{T}_*^{\text{fin}} - \mathbf{T}^0\right\|_2} \leq 1, \tag{5.26}$$

and the relative distance from the desired response as

$$D_i \equiv \frac{\left\|\mathbf{T}_i^{\text{fin}} - \mathbf{T}_*^{\text{fin}}\right\|_2}{\left\|\mathbf{T}_*^{\text{fin}} - \mathbf{T}^0\right\|_2}. \tag{5.27}$$

For convenience we introduce the response specificity μ as

$$0 \leq \mu_i \equiv (1 + D_i)^{-1} \leq 1 \tag{5.28}$$

such that the desired response has relative magnitude and specificity one, $\Delta_* = \mu_* = 1$.

The results are shown in Fig. 5.7, in the form of joint frequency histograms of magnitude Δ and specificity μ, $p(\Delta, \mu)$ in 5.7 (b,e). Their marginals, $p(\Delta)$ and $p(\mu)$ are shown in Figs. 5.7, for networks and proteins respectively in panels (a,c) and (d,f).

The histograms are evaluated for Δ_i and μ_i for all i surface pairs and triplets of beads on the surface of trained networks and for all i binding pockets in proteins. All values are normalized with respect to the given network, and the statistics are combined for all

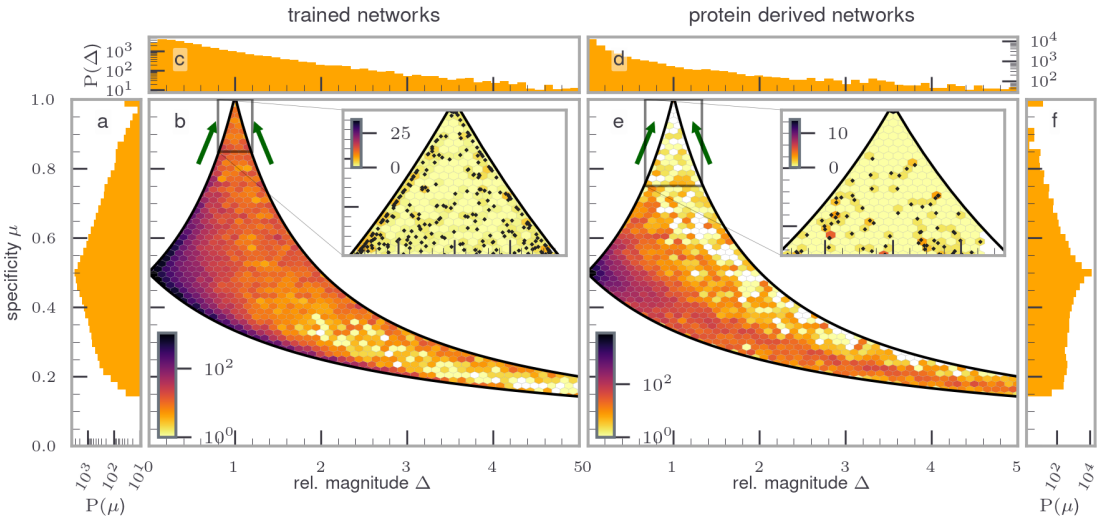


Figure 5.7: Identifying optimal source pockets for a specific allosteric response. Joint frequency histograms of magnitude Δ and specificity μ, $p(\Delta, \mu)$ (b,e), as defined in Eqs. 5.26 and 5.28 respectively, and their marginals, $p(\Delta)$ (c,d) and $p(\mu)$ (c,d), for networks and proteins, respectively. The black lines denote the bounds imposed by the three triangle inequalities between the norms in Eq. (5.29) and the green arrows point towards the desired response at $(1, 1)$. A magnification of the desired region is shown in the insets, dots indicate individual data points.

networks and proteins separately. The region into which all responses fall is described by the triangle inequalities between the norms[6] in Eqs. 5.26 and 5.28, which read

$$\begin{aligned} \mu(\Delta_i) &\geq \frac{1}{2+\Delta_i} \\ \mu(\Delta_i) &\leq \frac{1}{\Delta_i} \\ \mu(\Delta_i) &\leq \frac{1}{2-\Delta_i}. \end{aligned} \tag{5.29}$$

The corresponding bounds are depicted with black lines in Fig. 5.7.

Most of the candidate pockets show almost no response $\Delta \approx 0$, ranked at around $\mu \approx 1/2$. Note that this is in contrast to the observations from [Yan et al., 2018] (SI) who find that *a large response at the active site can be triggered by binding anywhere* for networks designed to propagate displacements, and is far more in line with what is expected for proteins. There is a considerable amount of pockets which lead upon binding to a strong response, that is however not in the desired direction caused by the real source ($\Delta > 0$ and $\mu < 1/2$),

[6]Note that the metric properties after the rotation and translation are maintained, a proof is given in the Appendix of [Fogolari et al., 2012].

and is thus only of secondary interest for drug design. Just very few pockets cause a strong and specific response, $\Delta \approx 1$ and $\mu \approx 1$ as shown in the inset in Fig. 5.7. The trained networks and proteins display qualitatively similar behavior, except that in proteins fewer pocket candidates lead to a response that is close to the desired one, possibly caused by a stronger specificity due to stronger evolution with additional constraints that we did not account for in our training.

Depending on the desired response in a practical application an optimal result will probably not be reached. In this case source candidates closest to the optimum have to be considered. A fallback would be to consider also pockets that only lead to a strong but nonspecific response, as these still carry allosteric information.

Chapter 6

Propagation and Prediction

As we have seen at the end of Chapter 2, the exact role of normal modes, and especially the soft modes, in the allosteric response is currently not clear. There is ample evidence showing that in many systems there is a good overlap between a soft mode or a subset of soft modes and the allosteric response. However, this can not be considered as a general principle because there are many counter-arguments and counterexamples demonstrating the failure of describing complex motion with eigenmodes stemming from the first mode in the first energy basin. One must account at least for linear combinations of multiple modes. It may even be that the modes do not play a dominant and direct role in determining the allosteric response. For the specific rearrangement of the target site, which is essential for precise control, in general a balanced interplay of many collective modes derived from multiple expansions along the constrained minimum path might be required.

Additionally, breaking down allosteric motion into single soft modes, such as hinges or shear mechanisms, has recently been revealed to be an impossible endeavor, as there is no single mechanism underlying allostery; instead it is a continuum of mechanisms [Liu, 2021]. Therefore, we do not even attempt to translate the movement into one that is easily discernible to the human eye, but instead maintain the way of thought of linear combinations of modes, not necessarily limited to the original, unperturbed configuration.

We found in Section 5.5 that in general allosteric transmission is non-reciprocal, thus indicating that there must be a fundamental difference between the source and the target site.

In this chapter we develop an analysis that is independent of the number of modes and allows the investigation of the differences between allosteric source and target sites. We

find a criterion that distinguishes the allosteric source site from the other binding pockets and employ it to predict sources in real proteins.

6.1 Local softening of modes

Building upon the findings of non-reciprocity in our trained networks and the implications this has for directionality of the allosteric response, we now try to quantify the difference between the source and target site.

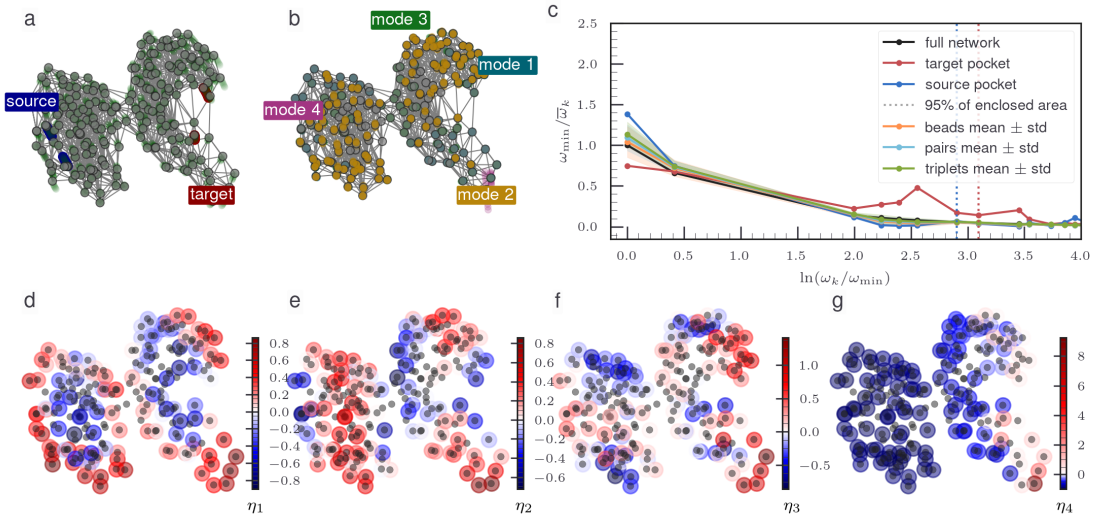
Spot-sampled investigations of the trained networks suggest that the soft modes, although often acting on the length scale of the whole network, tend to involve only the parts where the target pocket is located. The source pockets do not seem to participate in these soft modes at all. To quantify this, we calculate the projections of the relevant soft modes onto both pockets. Reminiscent of the collective springs, which the eigenvalues actually represent, we refer to small projections, expected at the allosteric sources sites, as *local hardening*, while large projections at the target sites are coined *local softening*.

We define the "local eigenvalues" as

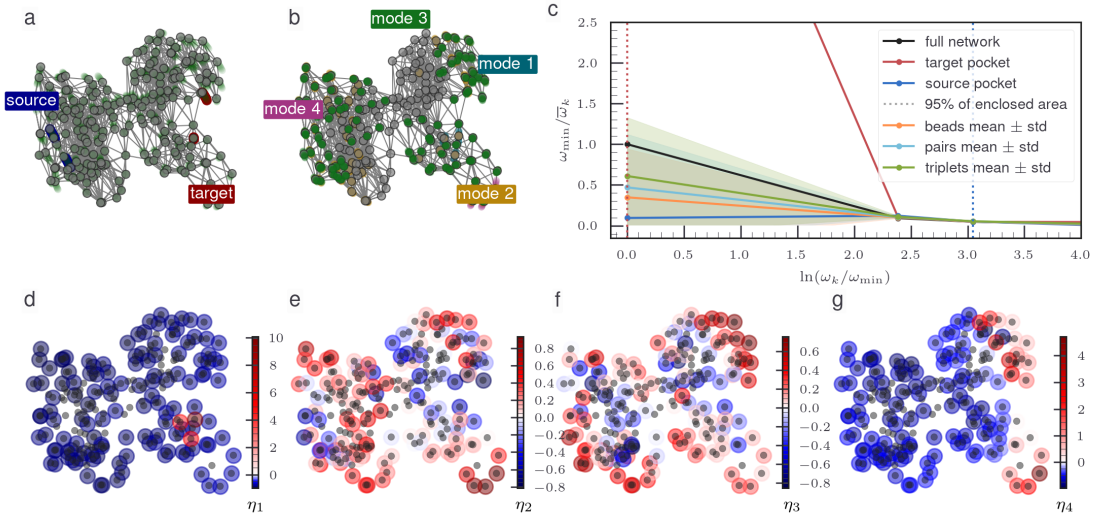
$$\bar{\lambda}_k(A) \equiv \sqrt{\frac{N_A}{N}} \frac{\lambda_k}{\left\| \hat{\mathrm{P}}_A \mathbf{v}_k \right\|_2} \equiv \frac{\lambda_k}{\chi_k(A)}, \tag{6.1}$$

where $\hat{\mathrm{P}}_A \boldsymbol{v}_k$ denotes the projection of $\boldsymbol{v}_k$ onto a relevant subset A of N_A beads, e. g. a set of source $A = S$ or target $A = T$ beads, and $|| \cdot ||_2$ the Euclidean norm. N is the total number of beads such that for the extremal case of projecting onto all beads $N_A = N$, where the projection $\hat{\mathrm{P}}_A \boldsymbol{v}_k = 1$ due to orthonormal eigenvectors. Moreover $\chi_k(A) = 1$, thereby recovering the bare spectrum.

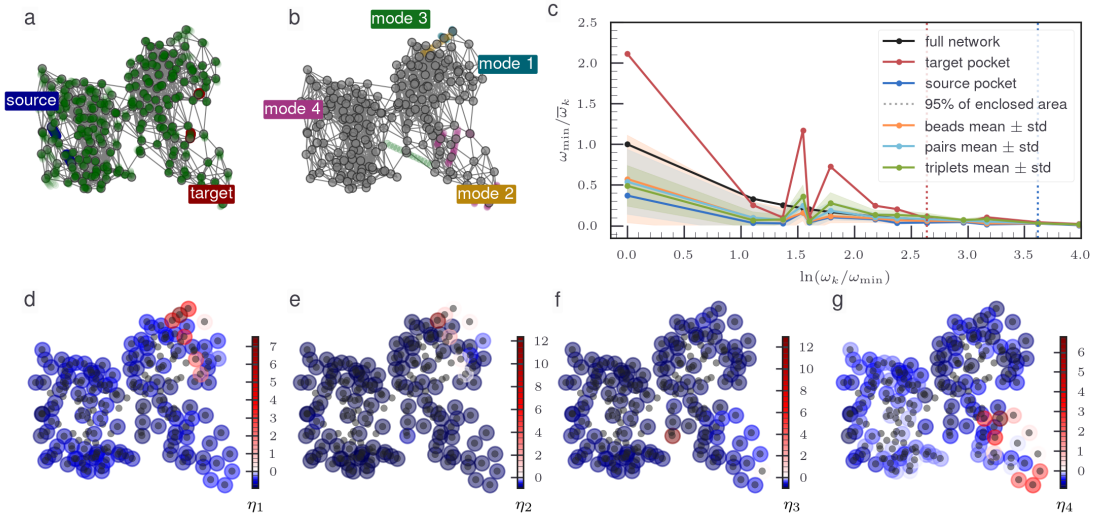
We introduce the local *softening* of mode k over the set of beads A $\eta_k(A) = \chi_k(A) - 1$ (which is equivalent to the *hardening* when $\eta_k(A) < 0$). The combined effect of the first M relevant (that is, they show a pronounced softening/hardening in the spectrum, as illustrated in panel (c) of Figs. 6.1 and 6.2) modes is described by the dimensionless effective softening $\eta(A) = \sum_{k=1}^{M} (\lambda_{\min}/\lambda_k)\, \eta_k(A)$, where M is the upper cutoff beyond which contributions (to the enclosed area under the respective spectra) are negligible and we neglect the six zero-modes (i. e. $\lambda_{\min} = \lambda_1 > 0$).



(a) initial random network

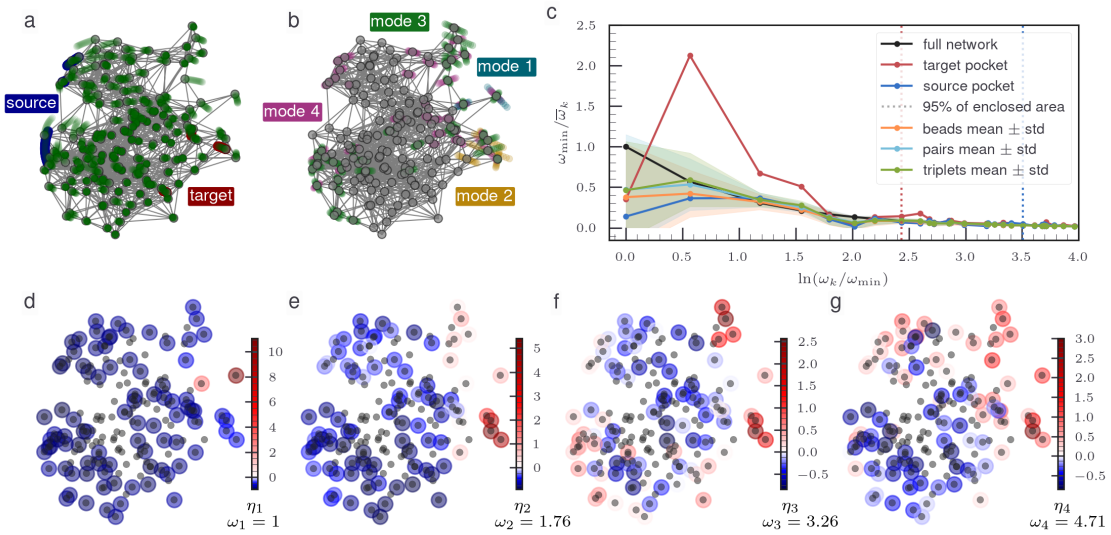


(b) trained variant: antisymmetric

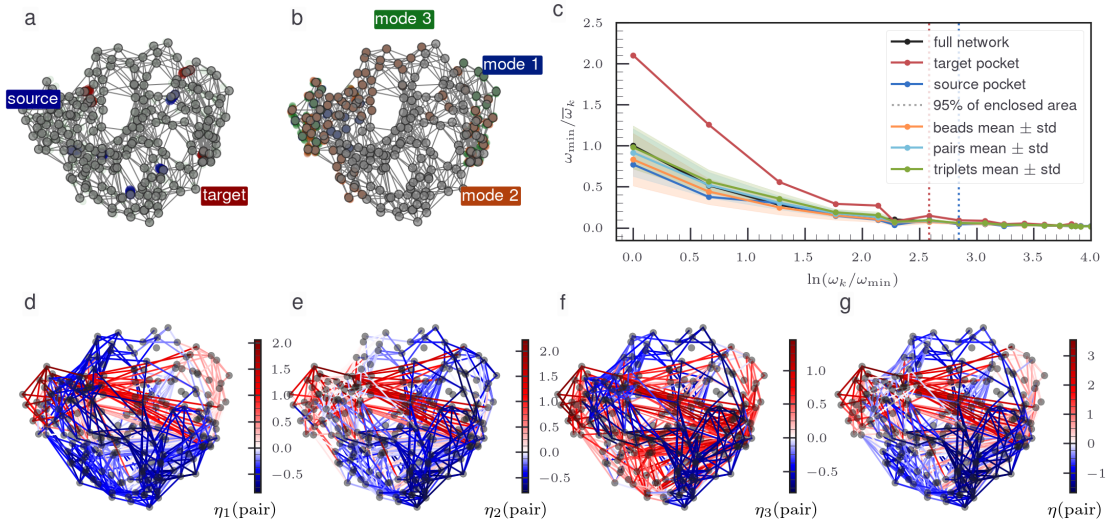


(c) trained variant: symmetric

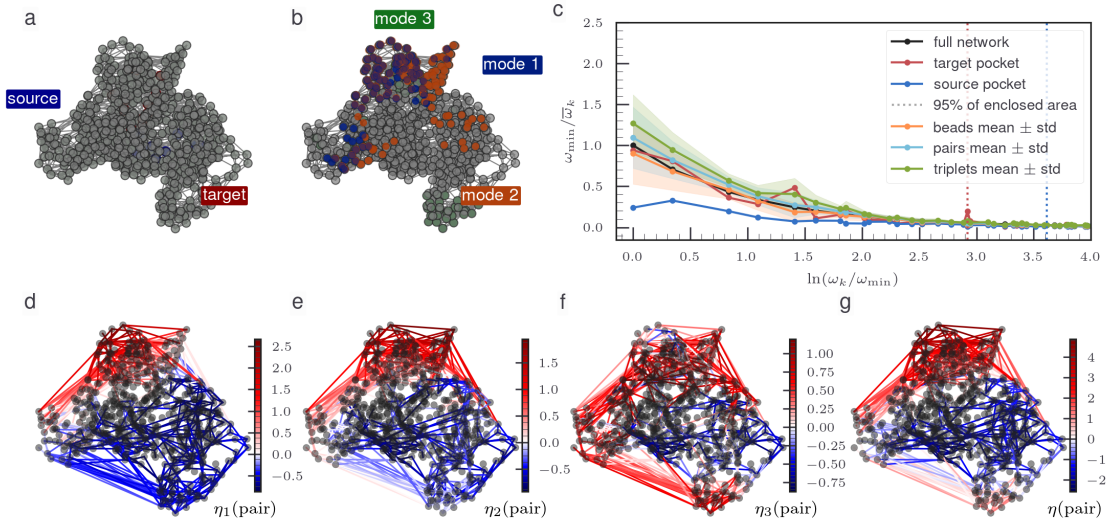
Figure 6.1: Local spectra for two trained networks and their ancestor, the initially random network [Flechsig, 2017].



(a) One of our trained networks



(b) Adenylate Kinase (ADK) (open, apo)



(c) Human Serum Albumin (HSA), (holo, with Warfarin bound)

Figure 6.2: Examples for the local spectra in three networks. Additional examples are shown in the Appendix A.8

Examples of the localization of the first few relevant modes in trained ENMs are given in Fig. 6.1 for structures available from the literature [Flechsig, 2017], and in Fig. 6.2a for a network we trained ourself. Localizations of soft modes in proteins are depicted in Fig. 6.2, for ADK (Fig. 6.2b) and HSA (Fig. 6.2c).

Figs. 6.1 and 6.2 share the same layout: Panel (a) shows the ENM and its response depicted in green, where the source perturbation is highlighted in blue and the target's response in red. Panel (b) depicts the first four non-zero eigenmodes of the system, with only those beads drawn that move substantially. The dimensionless mode-softness $\lambda_{\min}/\bar{\lambda}_k(A)$ for all k up to just above M is shown in panel c, for different sets A. High values in the reciprocal plot mean *soft* local modes and vice versa for *harder* modes. The source $A = S$ (blue) and target $A = T$ (red) pockets show strong deviations from the full spectrum $\lambda_{\min}/\lambda_k$ where $N_A = N$ (black). As anticipated, the soft, allosterically relevant modes show high projections at the target pockets, indicating that these modes' motions concentrate there. On the contrary, the beads of the source pockets experience rather small projections of the same modes. It is only in the untrained network that we see a qualitatively different behavior; in Fig. 6.1a the soft modes span the full network and no systematic differences between projections at different subsets are recognizable.

We also clearly observe that the fraction of modes relevant for the target pockets' motion can vary drastically between networks and proteins. The examples we studied only in rare cases allow to deduce that a single soft mode drives the allosteric response. Matching the results for the source and target projections nicely is the spectral behavior of the other proposed pockets in the networks. The averages $\pm$ standard deviation of these are shown for beads (orange), pairs (cyan) and triplets (green) on the surface of trained networks, and for triplets in the pockets of the proteins. Their lines follow the trend of the full inverse spectrum and the area that their variance covers in the reciprocal plot lies mostly in between the source and target pocket spectra.

The dashed lines in panel (c) depict the upper threshold M, up to which we consider modes to be relevant. These relevant modes represent the modes which enclose 95% of the area under the corresponding curves. Note that panel (c) is just a small cutout of the full spectrum, only the first $\sim$25 modes out of $3N$ are shown with N being on the order of up to a few hundred.

Panels (d-f) show the local hardening/softening of the four (three for proteins) lowest eigenmodes for all pockets, and the corresponding total hardness in (g) for proteins. A pronounced hardening in the vicinity of the source and a softening near the target beads is evident.

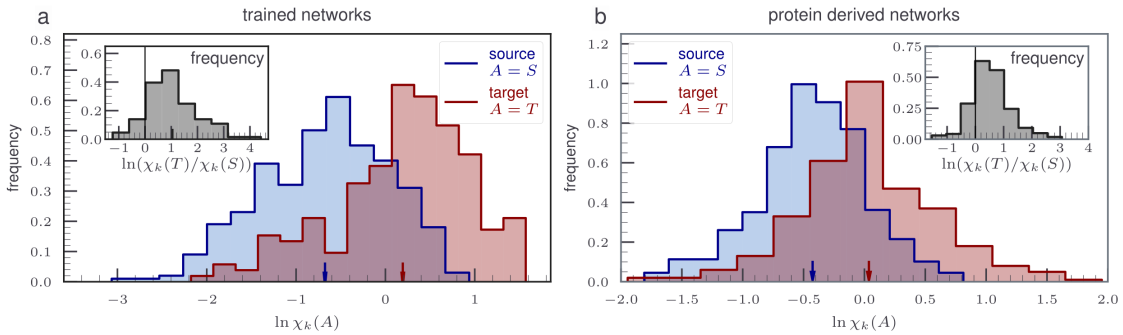


Figure 6.3: Frequency histograms of local softness/hardness of relevant modes, $\chi_k(A) = \lambda_k/\bar{\lambda}_k(A)$, for the source ($A = S$, blue) and target ($A = T$, red) pockets evaluated for all trained networks (a) and proteins (b). A two-sample Kolmogorov-Smirnov test [Smirnov, 1948] yields p-values 5.6×10^{-16} and 4.4×10^{-16} for networks and proteins, respectively, thus confirming that the distributions of χ_k for the source ($A = S$, blue) and target ($A = T$, red) pockets are indeed conclusively distinct.

If these findings are significant on a statistical level, we should in fact *typically* observe $\chi_k(S) < 1, \chi_k(T) > 1$, and $\chi_k(T) > \chi_k(S)$ for the relevant modes. Therefore we inspect the empirical distribution of local mode softness, χ_k, and relative target/source mode softness, $\chi_k(T)/\chi_k(S)$ (grey insets), over the set of all relevant modes (i. e. $k \leq M$) and all trained networks and proteins.

The frequency histograms (on a logarithmic χ_k-scale) are shown in figure 6.3 for trained networks (a) and proteins (b). Distributions for source (blue) and target (red) beads provide evidence that the relevant modes are indeed locally softer at the target and harder at the source sites, respectively.

The apparently big overlap between the distributions might at first seem to weaken the findings. However, if we study the distribution of the relative target/source mode softness $\chi_k(T)/\chi_k(S)$ we notice that per mode k, it holds that $\chi_k(T)/\chi_k(S) > 1$ (on the logarithmic scale: $\ln[\chi_k(T)/\chi_k(S)] > 0$). We see in the insets that this is almost always the case, which further supports the conclusion that a relevant mode is softer at the target site than it is at the source.

We can thus conclude that these results fully confirm our hypothesis, which was that allosteric source and target sites are fundamentally different. However, although this softness gradient already encodes the directionality of signal propagation in the networks—an input perturbation at the target site would simply be translated in a very local response of the direct neighbors and thus not propagated far—it is not precise enough to pin down exactly where the true source pocket lies. We see this in panels (d-g) of Figs. 6.1 and 6.2: the hard

and soft regions do not show pronounced peaks for the actual source and target pockets, they only classify their overall regions correctly. For accessing such specific information, a full-response scan as described in Section 5.6 seems to be unavoidable.

6.2 Mode mixing during response

The (iterative) linear response naturally consists of two steps, (i) the perturbation—constraining of the source beads $\delta\boldsymbol{c}$, and (ii) the response—relaxation of the rest of the network $\delta\boldsymbol{s}$. According to this partition we analyze the overlaps of the networks' motion with their eigenmodes, for both the self trained and the protein-derived networks.

Before we can take the scalar product between a $3N$ dimensional mode and a $3N - F$ or F dimensional displacement vector, we fill them up to full dimensionality with zeros, such that the assignment of nonzero entries and source/non-source beads is correct. All vectors are normalized. The projection is then

$$\alpha_k = \langle \delta\boldsymbol{r}_i | \boldsymbol{v}_k \rangle. \tag{6.2}$$

In order to make networks inter-comparable, the number of eigenmodes must be normalized. Of greatest interest is especially the difference of projections on the softer modes. Therefore we spread the eigenmode index for each network on a logarithmic scale after dividing by $3N$. The complete mapping for normalizing the mode indices is then $\log_{10}\left(\frac{9(k-6)}{3N}+1\right)$ for $k = 6 \ldots 3N$. We take the absolute of the square root for the projections $\sqrt{\alpha_k}$ and plot them on a logarithmic axis. The results are shown in panels (c) and (d) of Figs. 6.4 and 6.5 for the trained networks and the protein-derived networks, respectively.

The first part of our hypothesis, namely that the input couples to stiff modes and relaxes along soft modes, is convincingly confirmed. The two-dimensional histograms show a decline towards soft modes for the overlap of the eigenmodes with the displacement during constraining. Note the logarithmic scale; the shifts in the projections for the networks drop about one and the proteins about half an order of magnitude.

Far more pronounced, however, is the rise in the overlaps during the relaxing step towards softer modes. The change in magnitude is here almost three decades for the networks and two for the proteins. Apparently it is more important for the input displacement to *avoid the softest modes* than for the relaxation to couple only to higher modes in general. This can be deduced from the different ranges over which the projections onto low modes drop and rise, for the perturbation and the relaxation, respectively. This range covers the

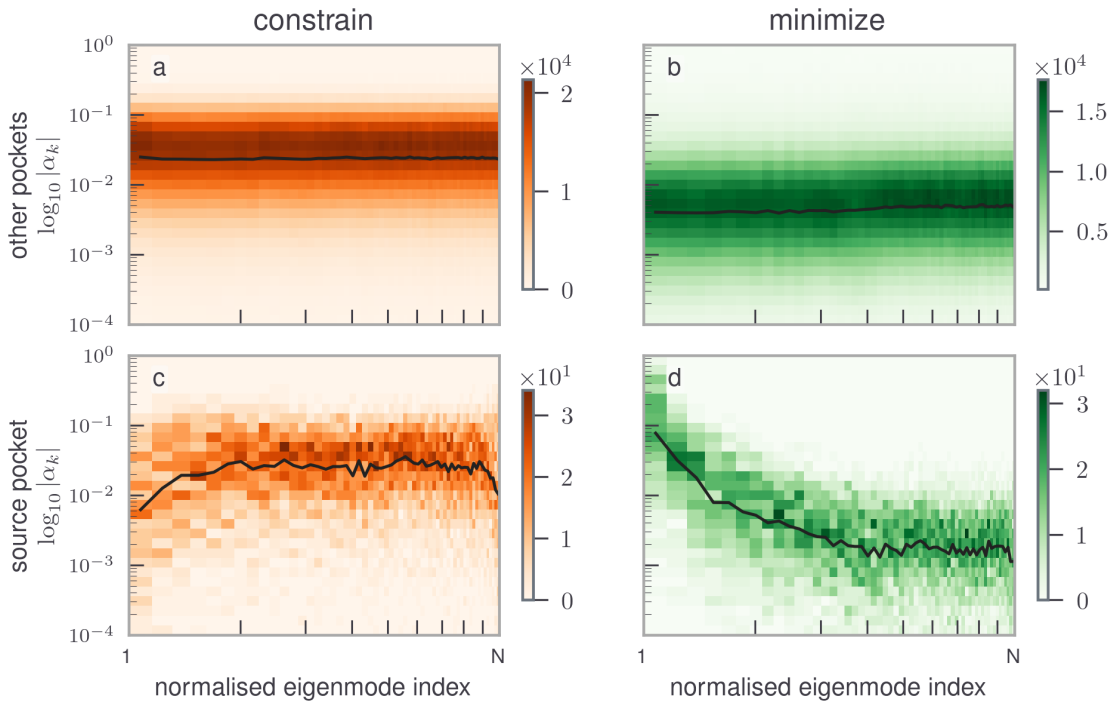


Figure 6.4: Projections of the perturbation (a,c) and the response (b,d) onto the eigenmodes of the Hessian for perturbations of all pockets (a,b) and that of the real source (c,d). Data shown for trained networks.

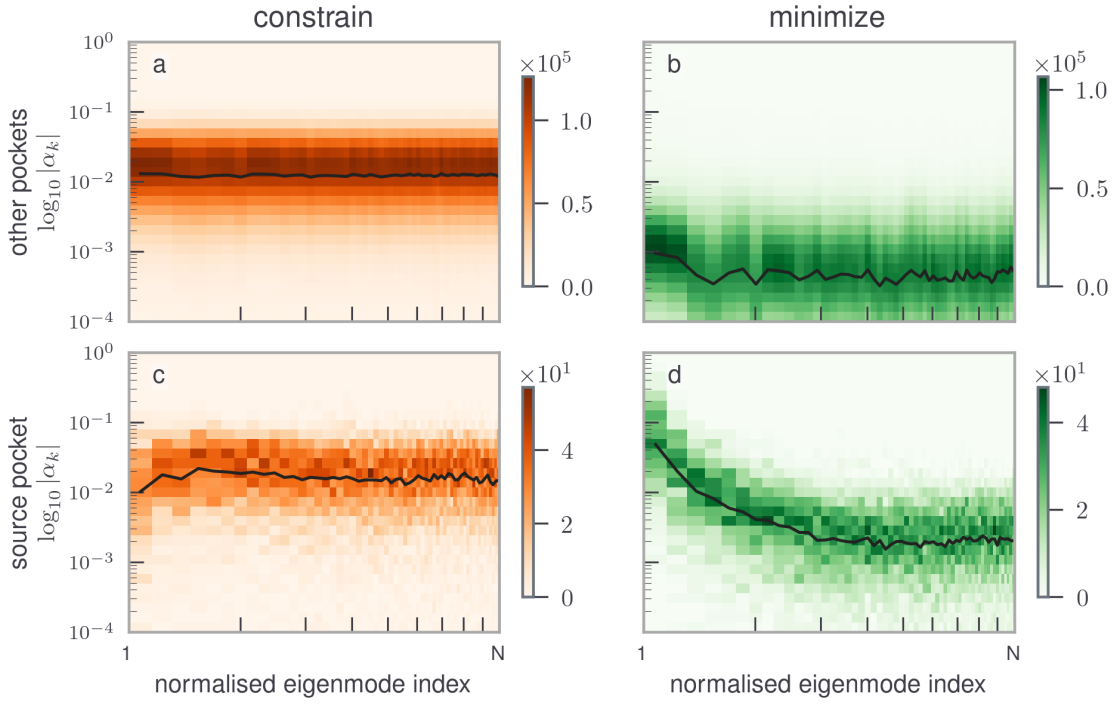


Figure 6.5: Description see Fig. 6.4. Data shown for protein-derived ENMs.

first two logarithmic mode bins during the relaxation step while the constraining step only shows pronounced low projections within the first bin. An interpretation for extremely high modes, condensed to the very right of the plots, is not possible. These are the motions that are on length-scales where the approximation of the ENM breaks down. However, the most pronounced difference between the mode coupling of constraining and relaxing lies at soft to intermediate stiff modes.

The relevance this has for the allosteric effect becomes especially clear when we compare the projections belonging to other input pockets, which are those not carrying an allosteric effect (see panels (a) and (b) of Fig. 6.4 and Fig. 6.5, respectively for the networks and proteins). There we carry out an analogous analysis for all other possible binding pockets, which we determined in Sections 4.2.4 and 4.1.2. The projections belonging to these pockets show almost no qualitative change over the modes, with an exception being the relaxation of the proteins, where we see a small jump towards higher modes at the first bins. The projections are still extremely small when compared to the relaxation of the real source beads.

These findings indicate that the response to closing random (aka non-allosteric pockets) is short ranged and dissipates directly in the vicinity of the pocket, caused by the nonspecific mixing with *both soft and stiff modes*. The concerted interplay that is needed to couple stiff to soft modes is specific for the real allosteric source pocket.

Now we can incorporate the fact that higher modes tend to localize at smaller length scales in the system and are located around regions with stronger stiffness, as described in Section 2.5. Although the catalytic motion at the active site of proteins seems to happen along the soft modes, these intermediate high modes are not unimportant for the dynamics of the protein. For long-range mechanical interactions in ENMs of allosteric proteins they indeed seem to be crucial.

These findings shift the paradigm of allosteric signal propagation in networks and proteins from a purely soft-mode based version towards a two-step thinking, which differentiates strongly between the allosteric in- and output. Nevertheless, these results do not *per se* require more information than is available in the first minimum; these overlaps are generated with a single step of the displacement based linear response algorithm. A quick scan over a $\sim$300 amino acid protein is performed in minutes on a standard personal computer. Only for a more detailed investigation of the motion in the target pocket a nonlinear approach must be employed, as we saw in Section 5.6.

Considering that normal modes are—as the name suggests—orthogonal to each other, interaction between modes may sound non-intuitive at first. As soon as only parts of the

eigenvectors are taken into account however, the orthogonality relation does not exist anymore.

Studies of allosteric structures and their modes which are performed only at equilibrium may therefore be restricted to insights that involve only part of the important modes.

Remembering first, that in our response algorithm the perturbation $\delta\boldsymbol{s} = \boldsymbol{s} - \boldsymbol{s}^{(0)}$ and the response of the system $\delta\boldsymbol{c} = \boldsymbol{c} - \boldsymbol{c}^{(0)}$ are related by

$$\delta\boldsymbol{s} = -\underline{\mathrm{S}}^{-1}\underline{\mathrm{B}}^{\mathsf{T}}\delta\boldsymbol{c} \tag{6.3}$$

and second that $\delta\boldsymbol{s}$ couples to the soft modes while $\delta\boldsymbol{c}$ couples to all but the soft modes, we can gain some insight into where the mixing between soft and stiff modes actually happens and what this means for the selectivity of the allosteric source pocket. $\underline{\mathrm{B}}^{\mathsf{T}}$ is a *thin* matrix ($\in \mathbb{R}^{3N-F\times F}$) that takes the input $\delta\boldsymbol{c}$ ($\in \mathbb{R}^{F}$) and distributes it to the rest of the $3N - F$ beads. Thus $\underline{\mathrm{B}}^{\mathsf{T}}$ must contain entries which allow for the real allosteric input $\delta\boldsymbol{c}^{\text{allo}}$ to selectively avoid loading the wrong (soft) springs. The relaxation over the inverse $\underline{\mathrm{S}}^{-1}$ allows then to receive the loaded springs and—again selective for the allosteric input—transfer their energy into motion along softer modes. The orthogonality condition is thus bypassed by switching to the lower dimensional vectors $\delta\boldsymbol{s}$ and $\delta\boldsymbol{c}$ and the mixing occurs via the matrices $\underline{\mathrm{S}}^{-1}\,\underline{\mathrm{B}}^{\mathsf{T}}$.

The idea of an energy flow between modes is not entirely new, in the context of comparisons to modes in crystals such behavior has been proposed for proteins [Bahar et al., 1998] but was not pursued further. Our findings are also consistent with the reported interaction between the nucleotide binding pocket of myosin and the rest of the protein if one interprets the arising suppression of motions in the pocket in the presence of coupling between both subsystems [Zheng and Brooks, 2005b]. Even the puzzling role of a specific set of residues in hemoglobin, as indicated in Section 5.5, which are known *not* to participate in the global soft mode [Perahia and Mouawad, 1995] but react sensitively to the binding of O_2 [Xu et al., 2003], can potentially be explained when accounting for a mixing of the locally available stiff modes with the global soft mode.

6.3 Efficient energy transfer

The second part of the hypothesis was related to the energy-uptake at the allosteric site and the transfer of this energy towards the active site. We postulated that allosteric networks are designed in a way that they provide the collective lever which specifically transmits the

energy available from an input displacement at the real source towards the target pocket. Thus, we expect a high energy change for *both, constraining and relaxation*. It is straightforward to test this prediction, by analyzing the change in energy related to the two steps of the response.

The energy change during the constraining step is

$$\delta U(\delta \boldsymbol{c}) = \frac{1}{2}\,\delta \boldsymbol{c}^{\mathsf{T}} \underline{\mathrm{C}} \delta \boldsymbol{c}, \tag{6.4}$$

and during relaxation step (with $\delta \boldsymbol{r}$ split in $\delta \boldsymbol{c}$ and $\delta \boldsymbol{s}$, and $\underline{\mathrm{H}}$ split in $\underline{\mathrm{S}}$, $\underline{\mathrm{B}}$, $\underline{\mathrm{B}}^{\mathsf{T}}$ and $\underline{\mathrm{C}}$)

$$\begin{aligned} \delta U(\delta \boldsymbol{s}) &= \frac{1}{2}\,\delta \boldsymbol{r}^{\mathsf{T}} \underline{\mathrm{H}} \delta \boldsymbol{r} - \frac{1}{2}\,\delta \boldsymbol{c}^{\mathsf{T}} \underline{\mathrm{C}} \delta \boldsymbol{c} \\ &= \frac{1}{2}\,\delta \boldsymbol{s}^{\mathsf{T}} \underline{\mathrm{S}} \delta \boldsymbol{s} + \delta \boldsymbol{s}^{\mathsf{T}} \underline{\mathrm{B}}^{\mathsf{T}} \delta \boldsymbol{c}\ , \\ &= -\frac{1}{2}\,\delta \boldsymbol{c}^{\mathsf{T}} \underline{\mathrm{B}} \underline{\mathrm{S}}^{-1} \underline{\mathrm{B}}^{\mathsf{T}} \delta \boldsymbol{s} \end{aligned} \tag{6.5}$$

where in the last step we used the definition of the linear response to replace $\delta \boldsymbol{s}$.

For a given input displacement $\delta \boldsymbol{c}$ an allosteric network should fulfill:

$$\sup_{\underline{\mathrm{C}}} \left\{ \delta \boldsymbol{c}^{\mathsf{T}} \underline{\mathrm{C}} \delta \boldsymbol{c} \right\} \qquad \wedge \qquad \inf_{\underline{\mathrm{C}},\underline{\mathrm{B}},\underline{\mathrm{S}}} \left\{ \delta \boldsymbol{c}^{\mathsf{T}} \left(\underline{\mathrm{C}} - \underline{\mathrm{B}} \underline{\mathrm{S}}^{-1} \underline{\mathrm{B}}^{\mathsf{T}} \right) \delta \boldsymbol{c} \right\}.$$

Again we compare between possible pocket candidates of the structures and the known allosteric binding pocket, for both protein and trained ENMs. The comparison between source candidates and the real source is carried out via the ratio of the respective change in energy, $\delta U^{A}/\delta U^{\text{source}}$ where A is the set of beads belonging to the tested pocket. To keep the test perturbation comparable between the different scanned pockets, we always pull each pockets' beads towards their center of mass until their radius of gyration is decreased by a specific amount which is fixed by the structure. Specific values are given in Tables A.1 and A.2. In order to rule out trivial effects caused by different pocket sizes we divide the energy by the number of beads we are pulling, normalizing the energy change to a "per-bead" energy change.

Figs. 6.6 and 6.7 show histograms over the "per-bead" energy change for the constraining (orange) and the minimization (green) step, combined for all artificial allosteric networks. The energy changes caused by closing the real source pocket are indicated with dashed lines at ± 1. We can clearly observe that there is a significant trend for the relative energy changes of the non-allosteric binding pockets towards smaller values. The effect is not

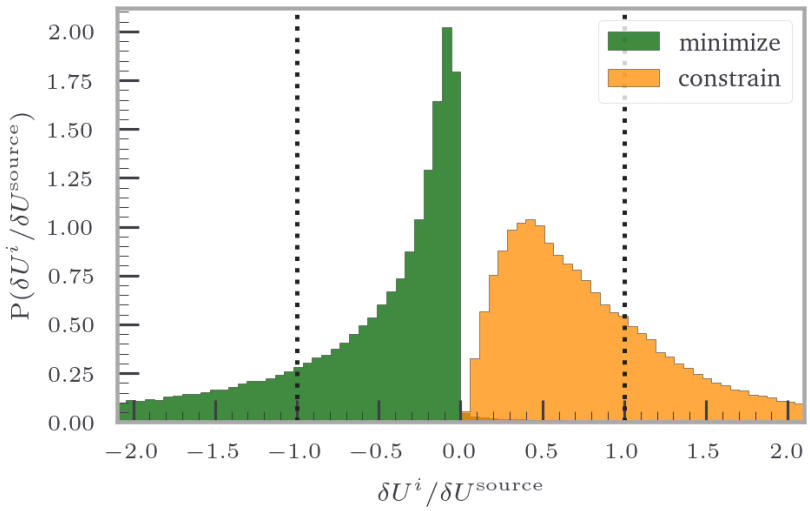


Figure 6.6: Histogram of energy changes for trained networks

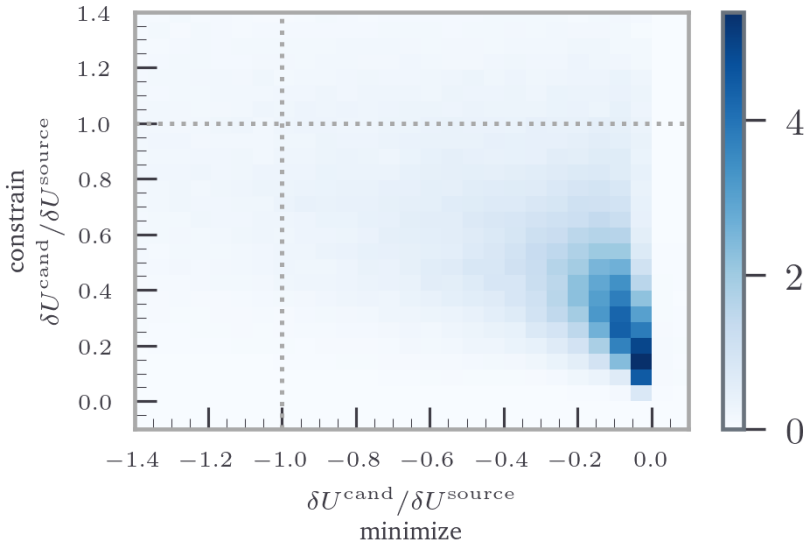


Figure 6.7: 2D Histogram of energy changes for trained networks

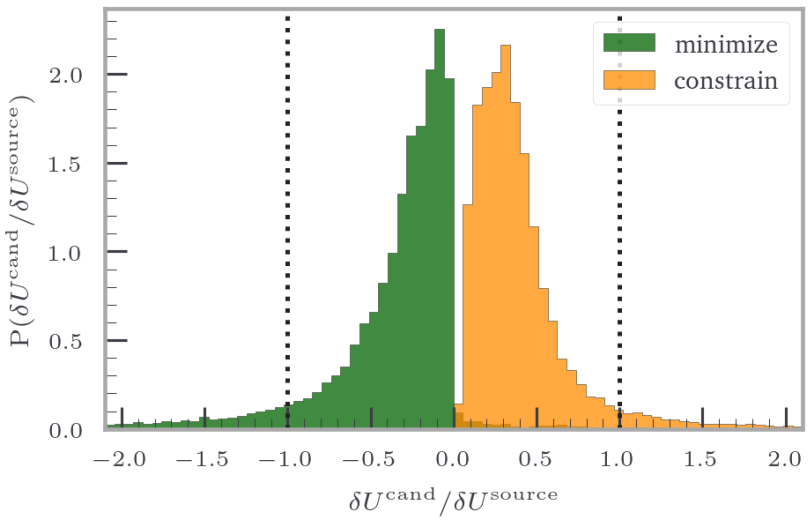


Figure 6.8: Histogram of energy changes for protein-derived networks

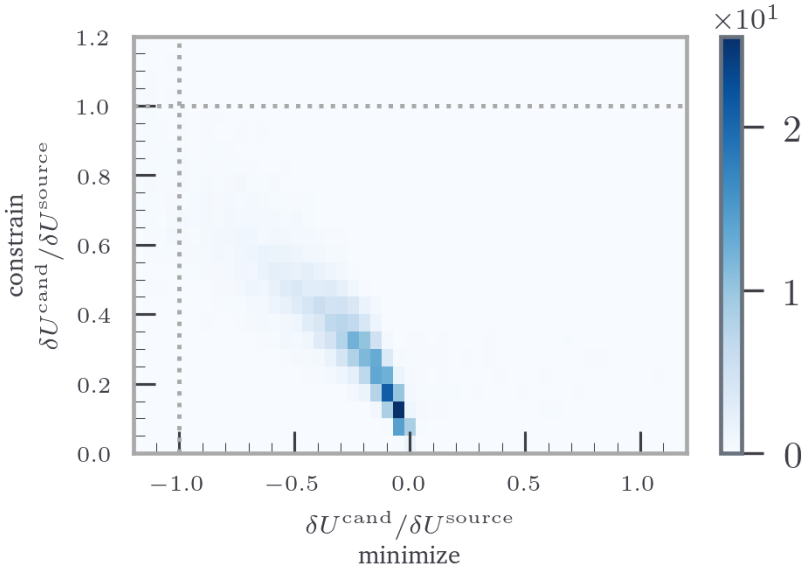


Figure 6.9: 2D Histogram of energy changes for for protein-derived networks

as strong for the constraining step as it is for the relaxation step. The long tails of the distributions, especially for the constraining step, indicate that there are also other beads that can induce displacements to the networks that lead to large energy changes. However, it is not guaranteed (for the trained networks) that the high energy caused by the input displacement also allows for a strong relaxation, as the relaxation distribution has a far less pronounced tail. The observation becomes evident from the shape of the distribution in the two-dimensional histogram in Fig. 6.7. We confirm, as anticipated in the previous section, that a high energy input is not a sufficient but only a necessary condition for allostery. The strong relaxation on the contrary has already as a prerequisite a high energy rise during the constraining step, rendering it a sufficient condition for an allosteric source, as we will see in Section 6.6.

Figs. 6.8 and 6.9 show the equivalent analysis for the protein-derived networks. The same effect is observable here, but is far more pronounced. The distributions are essentially symmetric, meaning that most of the binding pocket candidates do neither introduce nor transmit much energy through the network. The two dimensional histogram in Fig. 6.9 confirms this; there is almost a linear relation between input and relaxation. A possible interpretation for the superiority of proteins might be caused by the better training over longer timescales and with stronger selectivity of natural evolution.

Taken together with the results of the last section, these findings fully confirm our hypothesis that the coupling of stiff to soft modes allows allostery to efficiently propagate input displacements from the source to the target site.

Conditioning on μ and Δ

Fig. 6.10 shows the effect that conditioning the responses on either high Δ or μ have, similar as in Section 5.6. For all histograms we see a shift toward higher absolute values of the energy change, confirming that these high-energy pocket candidates indeed propagate allosteric signals specifically to the target site.

6.3.1 Subsample source pocket

To eliminate any further concerns, we take a closer look at the source pocket. We have indeed not scanned the large pockets found in Section 4.1.2 as a whole, but generated within them various combinations of how a ligand may bind there. The real source pocket is somewhat larger than these combinations, see Fig. A.7. Throughout we have always divided by the number of perturbed beads, their actual number thus should have no effect.

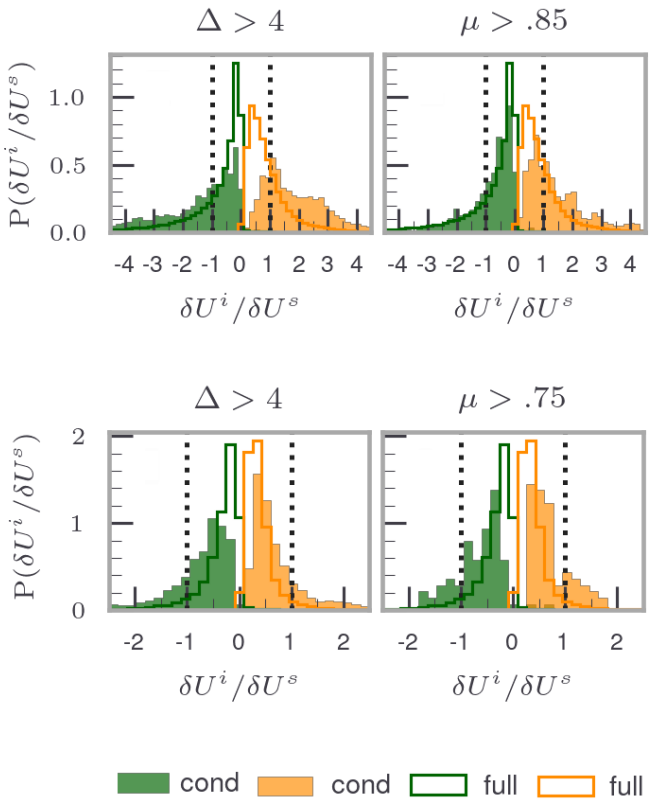


Figure 6.10: Conditional probability histograms of the energy changes during loading and relaxation for high Δ (left column) and μ (right column) shift towards higher values for both trained networks (upper row) and protein-derived networks (lower row), respectively.

However, a further control is advisable to check whether other possible effects caused by the pocket size differences do occur. Therefore we inspect the source pocket again by scanning across all combinations of triplets possible with the pocket' beads, equivalent to the sampling performed within the larger binding pocket candidates.

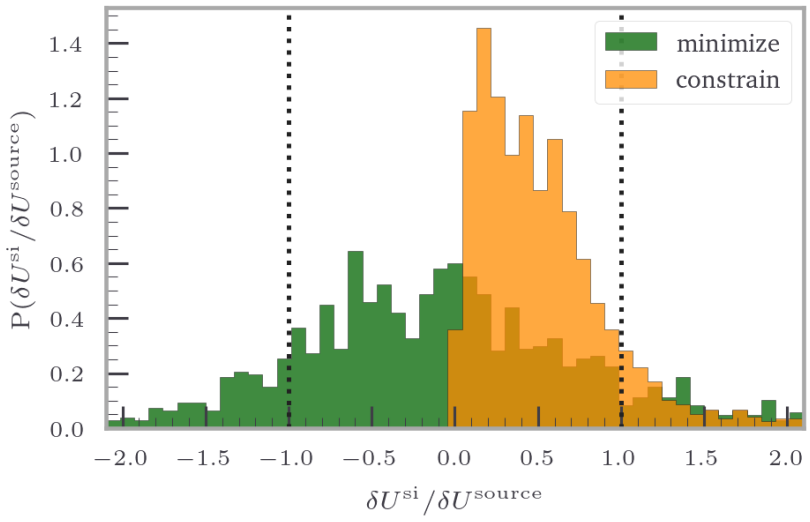


Figure 6.11: Histogram of energy changes for protein-derived networks (source subsampled)

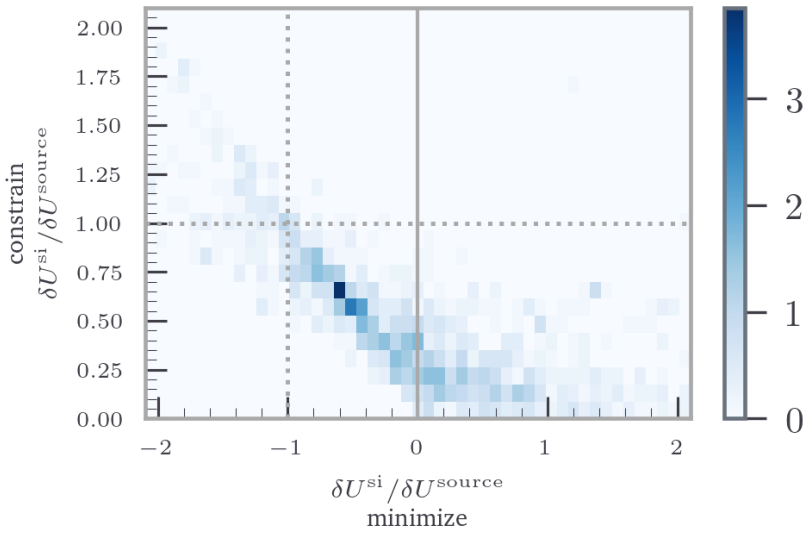


Figure 6.12: 2D Histogram of energy changes for for protein-derived networks (source subsampled)

In general, just because the source beads are found to bind to the ligand, that does not mean that all of them are equally important for transmitting the allosteric signal. It could also be that only a subset of them is actually responsible for loading stiff collective springs. Along these lines we can also interpret the results from the sub-sampling of the source pocket shown in Fig. 6.11 and 6.12. We see a significant difference with respect to the other pockets in the distributions of the energy change for the relaxation step and recover the known trend for the energy input during constraining. This means that indeed not all of the source beads contribute equally. Rather there exist critically important beads within the source pockets which are accountable for most of the energy input. If we do not pull these critical beads, but the other source pocket beads, which are however connected to the critical ones, the energy does not increase immediately during the constraining step, but only afterwards in the minimization step, that is, when the critical beads follow. In this respect, the (relaxation) movements of the critical beads compensate for the absence of energy input from the constrained, non-critical beads in the source pocket.

6.3.2 Large pockets

Another complementary check would be to directly perturb the large pockets and study the energy changes caused by them. We find that there is no qualitative difference between scanning smaller parts of the possible binding pockets and the whole binding pockets, as shown in Figs. 6.13 and 6.14. The exceptionally high values for energy uptake and release when pulling the real source pocket in contrast to almost all other pockets is still more than pronounced. This excludes a possibly trivial pocket-size effect.

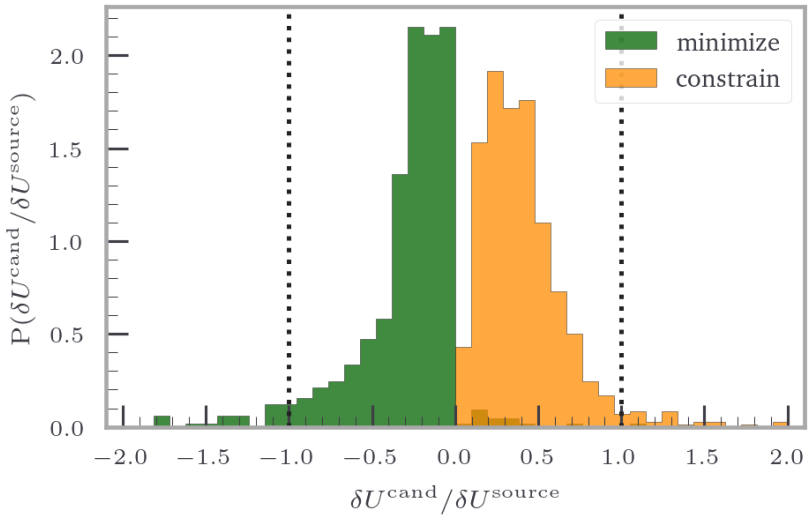


Figure 6.13: Histogram of energy changes for protein-derived networks (large pockets)

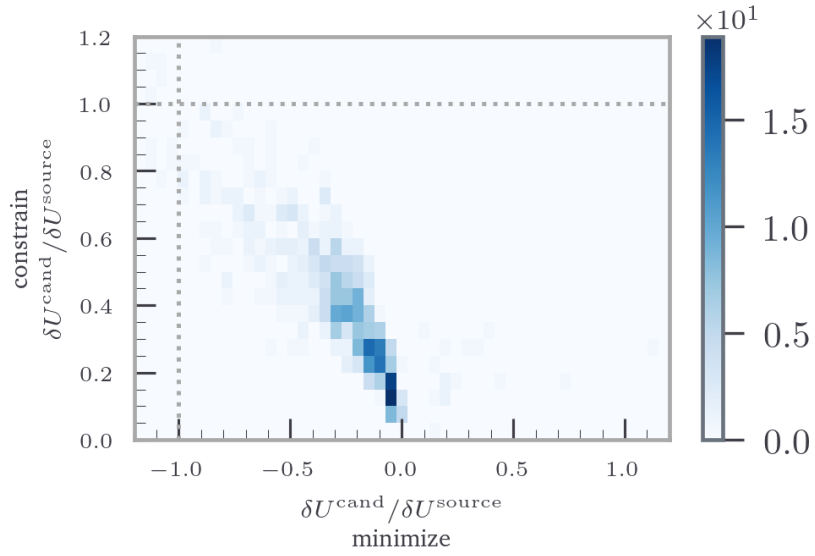


Figure 6.14: 2D Histogram of energy changes for for protein-derived networks (large pockets)

6.4 Illustrative example

To get an impression of the spatial distribution of the energy changes induced by the perturbation and subsequent relaxation for different pocket candidates in a network, we will take a protein-derived network as an example. Fig. 6.15 depicts these observables in Human Serum Albumin (HSA) (PDB ID 2BXD [Ghuman et al., 2005]), a monomeric protein. Panel (a) shows a cartoon representation of the protein with geometrically determined (see Section 4.1.2) binding pocket candidates, shown in color are the largest 9 out of a total of 77 analyzed pockets. The elastic network derived from the protein structure is shown in panel (b), coarse grained on a residue level, the beads are located at α-carbon positions. Colored beads constitute the potential binding pockets matching panel (a).

The input perturbation at the allosteric site (source pocket, blue beads) and the network's nonlinear response (green trace) is shown in panel (c), with beads at the active site (target pocket) highlighted in red. The experimentally known binding sites for the allosteric ligand (warfarin) and the regulated ligand (heme) constitute the source and target pockets, respectively (Information about the binding sites' residues can be obtained from [Ascenzi and Fasano, 2010]).

The fist four softest (nonzero) eigenmodes of the Hessian matrix of the network in panel (b) are superimposed with the structure in panel (d), only substantial motions are shown. Notice the lack of soft-mode participation of the source pocket beads. The impact of contractions of possible binding pockets on the potential energy of the network, for loading the springs around the source pocket in panel (f) and relaxing the network in panel (e), colors indicate the magnitude of the change in potential energy relative to closing of the real source pocket. The statistics over (e) and (f), respectively in green and orange histograms (analogous to Fig. 6.8) in panel (f), again we see that only a few pockets show a substantial change in energy.

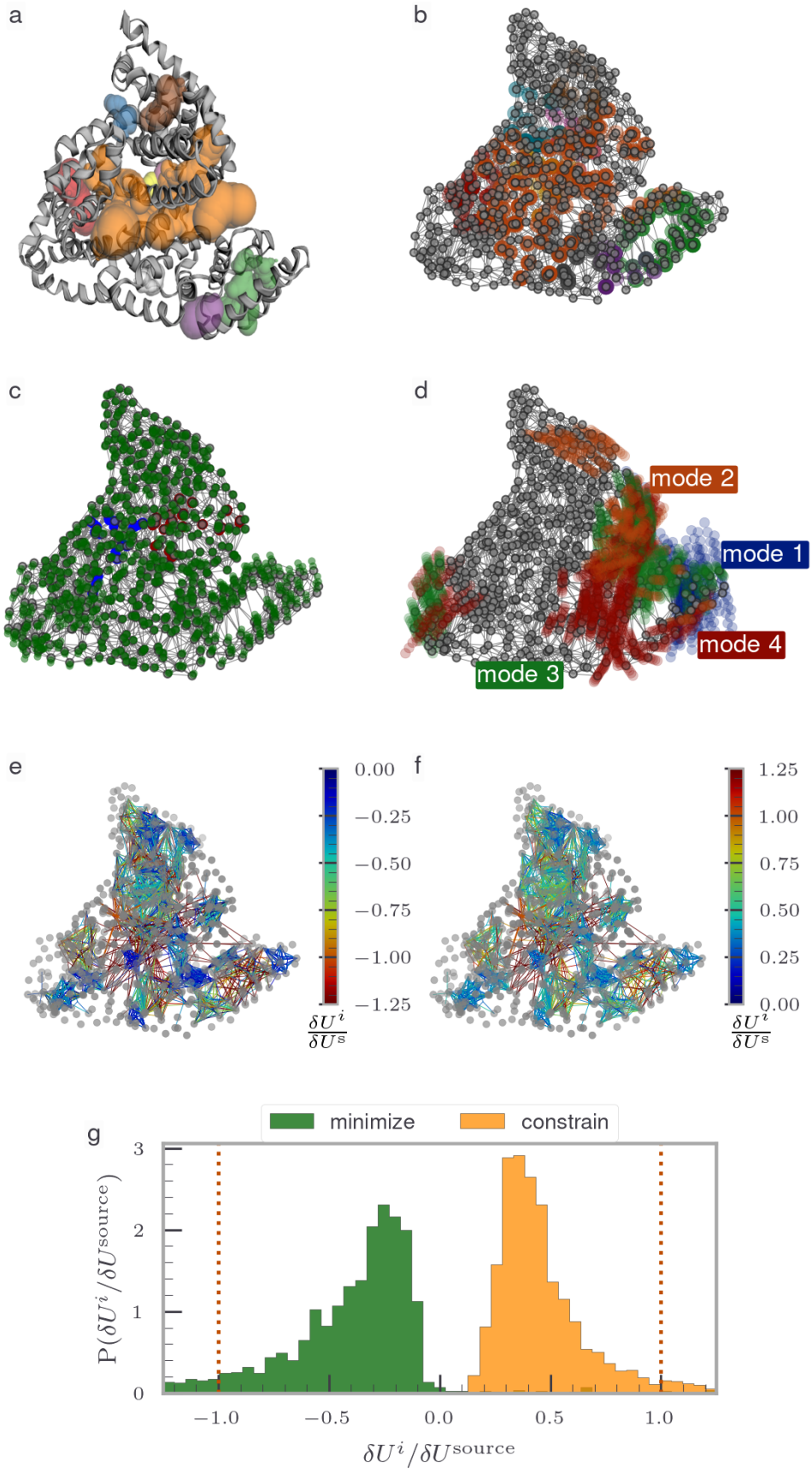


Figure 6.15: Example depicting the energy changes in a protein-derived network. For a detailed description refer to the text.

6.5 Interpretation

Our approach differs systematically from previous attempts to understand allostery in mechanical networks. Splitting the motion into perturbation and response allows us to draw a connection between previously unrelated facts. The conjecture that virtually all proteins exhibit allostery [Gunasekaran et al., 2004] and the fact that it is remarkably tractable to assign allostery into artificial[1] networks [Rocks et al., 2017] raises the questions about their common properties and why allostery is so ubiquitous in a wide variety of networks.

It has long been known that proteins are very similar in certain properties: They exhibit a similar vibrational spectrum from low to intermediate frequencies [Ben-Avraham, 1993]; the associated soft modes, which propagate over the entire length scale, tend to be relevant to protein function [Nicolay and Sanejouand, 2006, Na et al., 2018]; harder modes typically localize at smaller length scales near more rigidly connected structures [Bartlett et al., 2002, Yuan et al., 2003] due to the presence of inhomogeneities [Anderson, 1978, Anderson, 1958].

These are also the properties in which artificial networks and proteins are similar. We also saw that in both a specific allosteric response emerges from an elaborate interplay of harder modes that transmits input displacements from the allosteric source pocket to the active target pocket through the soft modes.

The existence of soft global and stiffer local modes belongs to the intrinsic properties to inhomogeneous media like proteins. It is not far-fetched to assume that nature takes advantage of this. These ingredients apparently do not need to be mixed strongly for allosteric effects to arise. Due to the high number of stiff modes these are available throughout the networks for being coupled to the softer ones by means of small mutations. Thereby it is possible to impose directionality through coupling from stiff to soft, and incorporating specificity through selecting critical residues at the source pocket. This enables proteins to be controlled precisely, just as the cell demands.

[1] Even more impressive is the successful design of a new allosteric site into a protein [Zhang and Bishop, 2007].

6.6 Predictive power

Using the specific role of the source beads, it is possible to predict possible source pockets.

To do this, we require a means to rank source pocket candidates by how much energy the network releases during the relaxation step[2] after we perturbed these source pocket candidates' beads during the constraining step. The data we have, however, is on a pair or triplet (with a few exceptions also quartet and quintets) basis. There is no unique way to reduce this data to a bead basis. We try to provide the simplest mapping: We rank pocket candidates, which is easy to do from the individual proteins' data in Fig. 6.8, and append for each of the pockets the respective beads as an ordered list. If a bead is already in the list it is not appended again. This permits a complete ranking of all the sampled beads. The beads at the front of the list are now highly likely to be source beads, according to our prediction.

To generate a binary assignment, we now need to decide on a threshold that classifies beads into source beads and non-source beads. For a Receiver Operating Characteristic (ROC) analysis [Park et al., 2004], this threshold is traversed, and the predictions are subsequently evaluated. They fall into the classes of true and false positives which are exactly the coordinates of the ROC plot.

A perfect classifier would give only true positives and zero false positives, i. e. a step function in the ROC graph. A random selection would give a straight line with slope one half. Classifiers that lie above this straight line are better than a random guess and the closer the ROC curve is to the step function (easily measured in terms of the Area Under Receiver Operating Characteristic curve (AUROC)), the better the classifier performs.

The prediction of source beads based on their energy-release performs extremely well, as can be seen in Fig. 6.16. The ROC curves for all proteins are far better than a random guess and reach high values for the AUROC values, with all of them being larger than 0.7. Our method performs significantly better than a recent study with a similar aim, which is relying on the reciprocity of allosteric signal propagation [Tee et al., 2018]. This is surprising as there is remarkably little information entering the analysis. Essentially, only a single step of the linear response algorithm is needed to access this information.

[2] Or by how much energy they contribute to the network in the constraining step. The results are similar on average, but show a larger variance. See Section A.10 in the Appendix. As the strong relaxation during the relaxation already requires a strong loading during the constraining step, and thus includes information of both events, we expect the relaxation energies to be more relevant.

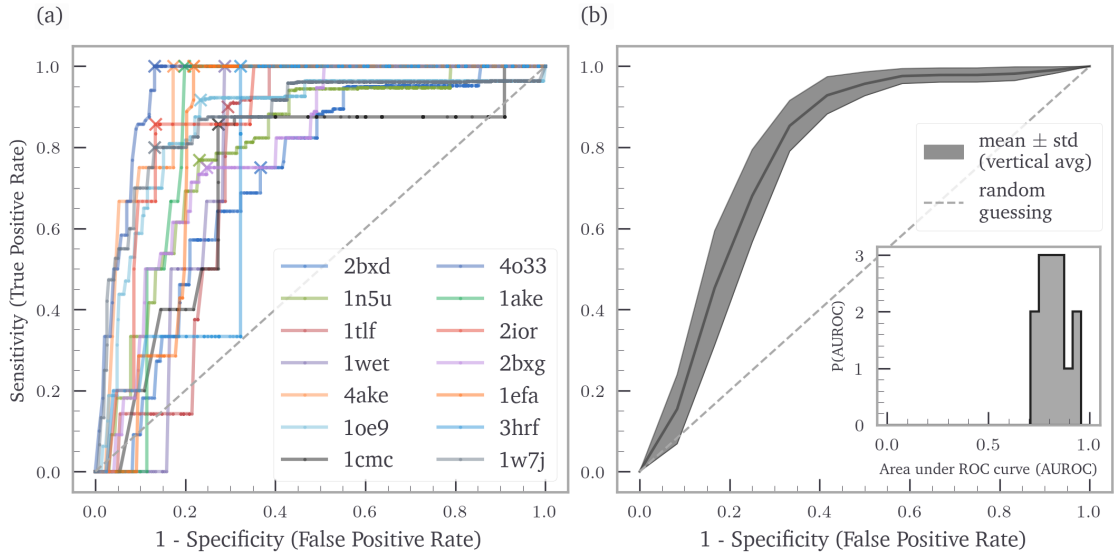


Figure 6.16: The Receiver Operating Characteristic (ROC) curve is an effective method of illustrating the performance of a binary classifier [Park et al., 2004]. (a) shows the curves for all proteins used in this work. Crosses mark the optimal threshold per protein. The grey shaded area in (b) shows the mean (vertical averaging [Fawcett, 2006]) of the ROC curves, ± standard deviation. The inset in (b) shows the Area Under Receiver Operating Characteristic curve (AUROC) values which are astonishingly high for a classification that relies only on a single initial structure per protein.

6.7 Scoring residue conservation

In evolutionary biology conserved sequences refer to identical or similar sequences found in proteins or nucleic acids (RNA and DNA) across different species. A conserved sequence is an indication that this sequence is selectively maintained; the interpretation is that this sequence is functionally relevant for the protein[3]

To check whether the residues that we predict to be crucially relevant for the allosteric action, we evaluate the degree of conservation of their corresponding part of the sequence. We first search for and retrieve amino acid sequences that are similar to each protein with Basic Local Alignment Search Tool (BLAST) [Ye et al., 2006], and subsequently employ *Clustal Omega* [Sievers and Higgins, 2014] to do a Multiple Sequence Alignment (MSA). In order to extract the data that is important for the respective protein of interest, we cut out the part from the MSA that corresponds to the sequence of this protein.

[3]This interpretation that has to be approached with caution, as there are also non-coding sequences on the DNA level that are conserved—at first sight without functional relevance [Asthana et al., 2007].

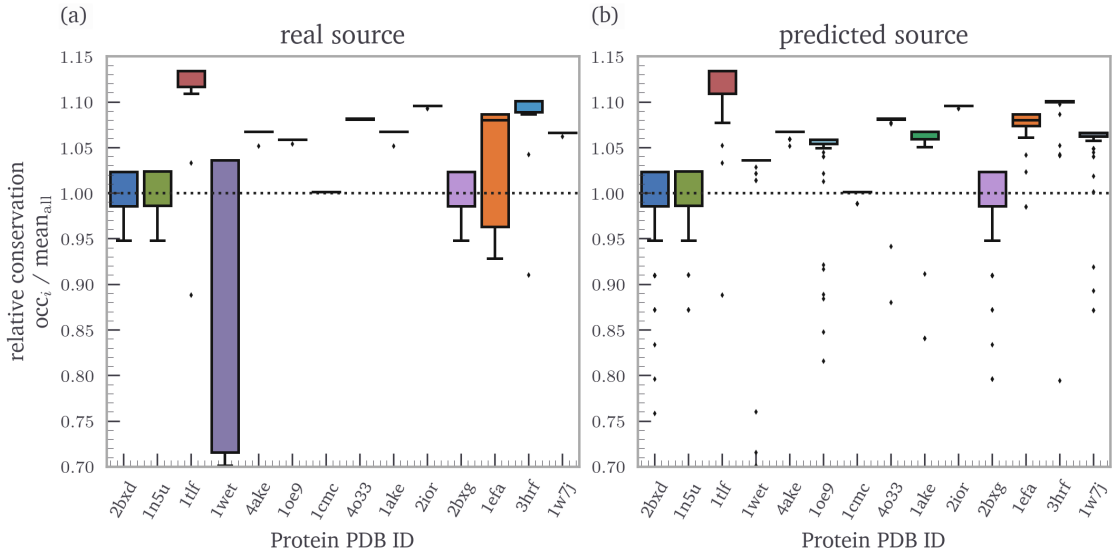


Figure 6.17: Boxplots illustrating the statistics over the mean degree of conservation per group of beads and proteins. The degree of conservation is given relative background mutation rate, i. e. to the average conservation of the full sequence. Notice that the predicted source beads are conserved with a higher rate than the real source beads. This could be related to the finding in Section 6.3 where we saw that not all source beads contribute equally to the energy input, some where drastically more important than others.

We then first calculate a background mutation rate which is the average degree of conservation of the full sequence, i. e. the ratio of how often the residue we have in our original sequence occurs at the same spot in the other sequences. Then we divide for each predicted source pocket residue its mutation rate by the background mutation rate; values larger than one correspond to a conservation that is more than average.

We find that the real source pocket residues are conserved more than the average, with a few exceptions. The beads we predict as source beads are conserved with the same or an even slightly higher rate. This may be interpreted as that not all of the residues in the source pocket are important for allostery although the ligand binds to them, especially when seen in context of the findings in Section 6.3, where we saw that not all of the source beads directly contribute to the energy input.

Chapter 7

Conclusion and outlook

7.1 Conclusion

The following quote nicely summarizes what we have achieved in this thesis; and it reminds us of the ultimate goal of drug design, towards which we have made an important first step.

> *Perhaps, as we learn more about allosteric systems it may become possible to develop algorithms to distinguish allosteric from active sites, and to design novel triggers to inhibit or activate proteins as required.* [Laskowski et al., 2009]

Indeed, we have shown that in order to understand the mechanical basis of allostery, it was essential to assume a fundamental difference between allosteric and active sites. This in turn allowed us to accurately predict the actual allosteric sites in proteins by analyzing solely the energy transfer in a single structure.

So, finally, we are able to answer quite precisely the questions we originally posed. First of all, the question of where the allosteric effect should appear has turned out to be only of secondary importance. On the one hand, often this particular effect and the actual site is already known in the proteins of interest which are to be controlled by drugs, and, on the other hand, it is much easier to predict, since the resolved structures often show distinct conformational transitions just there. Nevertheless, spectral analysis can still provide valuable clues here, even from a single structure, as the active site is predominantly located in the regions where the global soft modes exert their effects. A more precise prediction of these sites is not possible, since the effect of the soft modes is often extended over a larger area.

The input for the allosteric regulation is notoriously harder to predict as it does not participate in the well-studied and well-behaved soft eigenmodes. As the input couples to stiffer modes, it is impossible to predict a priori without a perturbation-based response which part of the protein is important—there are simply too many possibilities. However, as already indicated, the question of where to induce an allosteric effect can be answered with high precision. Excitingly, this is possible even without knowing the active location and works for *holo* (ligand-bound) structures as well as for *apo* (ligand-free) structures. A scan involving all possible binding sites of a protein, each of which undergoes a perturbation by contracting, allows the allosteric sites to be identified. The corresponding residues only have to be ranked according to how strongly the mechanical energy relaxes again after a perturbation.

The observations of the last chapter have confirmed our hypothesis that allosteric networks are exceptionally efficient at first receiving the mechanical energy associated with input deformations via stiff modes and then releasing it back to the targeted region via soft modes.

This also relates to the answer to the third question, about how the mechanical signal is propagated. Indeed, allosteric networks are optimized to accept displacements at a certain location particularly well and to load the collective springs in the vicinity of this site, whereupon the network converts the mechanical energy efficiently into long-range motion by coupling the local, stiff modes to soft, global modes. Allostery is thus at least to some extent a non-equilibrium phenomenon that only becomes fully apparent upon perturbation, assessments in equilibrium can, if at all, only account for half of the picture.

We have also seen that, depending on the system, the allosteric response can proceed either along a single mode or along a combination of several modes, and in some cases cannot even be described with the modes from the first energy minimum but may require a full nonlinear description. We have found the explanation for this in the rotation of the eigenmodes during the response itself, which varies strongly in magnitude between the studied networks.

As we have seen, the origin of long range coupling between binding sites in elastic networks is identical for protein-derived and artificial structures. Thus, the assumption that the essential properties of allostery are captured in the ENM is well justified. Last but not least the question arises, why the coarse-grained description actually works so well for describing allosteric motions, considering that the passage of barriers—usually regarded as essential—was not observed during the responses. The answer is found in the smoothening effect that accompanies every coarse-grained model[1]; local variations in the energy landscape are inherently flattened for residue level descriptions of proteins and especially ENMs

[1]This effect is depicted in Fig. A.9 in the Appendix.

are known to be capable of *sampling nearby configurations that are otherwise separated by local energy barriers* [Xu et al., 2003]. Our formulation for determining the response of proteins to perturbations now pushes the boundaries of what "nearby" refers to.

7.2 Outlook

The first straightforward application of the newly won knowledge would be to approach the evolutionary training of artificial networks from a spectral perspective. How exactly do the predicted observables (efficient energy transfer and mutual exclusive eigenmode projections) change during the training, and how, when and where does the selective coupling between stiff and soft modes emerge? This could also give the answer to the question why it is so remarkably easy for networks to adapt allosteric behavior [Rocks et al., 2017]; finally this could hint towards the reason for the ubiquity of allostery in proteins and other macromolecules. Coupled to this question is the naturally arising question of whether it is possible to pinpoint which modes exactly couple to each other and which residues are involved? The (probably nontrivial) answer must lie in the special structure of the two matrices $\underline{\mathrm{S}}$ and $\underline{\mathrm{B}}$, (in Eq. 5.8) as they connect between the input $\delta\boldsymbol{c}$ and the response $\delta\boldsymbol{s}$. Yet another extension related to the evolutionary training addresses entropic allostery. Is it possible to train for allostery without a conformational change [Cooper and Dryden, 1984], if we adopt a different cost function during training that addresses primarily the suppression of fluctuations?

Another interesting extension is to use the developed method to make quantitative and measurable predictions, by including the statistical mechanics description of the binding of ligands explicitly. Foundations for this work are developed in [Olaussen and Stell, 1991, Okazaki and Takada, 2008, Gilson et al., 1997], the computation of binding affinities and eventually *Hill-curves* [Hill, 1910] based on the configurations of the different allosteric states, should therefore be possible.

The underlying model, the ENM, could itself be extended in various ways to allow for more complex behavior, e. g. (partial) un- and refolding or the occupation of different, empirically known conformational states. The former might be incorporated by allowing either for an adaptive connectivity[2], or, with different types of springs, e. g. by introducing breakable springs, following [Poma et al., 2018]. Describing conformational motion covering multiple distinct energy basins is possible by iteratively combining ENM of different structures, e. g. the two allosteric states of proteins (assuming multiple experi-

[2]This is implemented in the software already.

mental structures area available), which goes along the lines of [Miyashita et al., 2003]. A possible implementation could linearly interpolate between the respective two ENMs by scaling their spring constants like it is e. g. done with *λ-dynamics* in MD simulations [Kong and Brooks III, 1996, Kohnke et al., 2020], where the protonation states of the titratable group can be varied continuously.

As discussed, the incorporation of implicit solvent models into the Hessian matrix is in principle possible. However, due to the computational overhead, another approach might be more promising. The covariance matrices of trajectories generated in MD simulations provide, upon diagonalization, access to eigenmodes that are similar to the normal modes of ENM. These so called principal components encode collective motions according to their variance and no harmonic approximation is implied. It is quite likely, that an equivalent analysis of these PCA modes confirms our findings with higher chemical detail.

However, the convincing success in predicting potential allosteric sites calls for an immediate application to computational drug design.

Acronyms

cryo-EM	cryo Electron Microscopy	15
NMR	Nuclear Magnetic Resonance	4
PCA	Principal Component Analysis	17
NMA	Normal Mode Analysis	8
ENM	Elastic Network Model	7
ANM	Anisotropic Network Model	19
GNM	Gaussian Network Model	19
PDB	Protein Data Bank	39
DNA	DeoxyriboNucleic Acid	2
RNA	RiboNucleic Acid	2
MSA	Multiple Sequence Alignment	6
MD	Molecular Dynamics	4
LRT	Linear Response Theory	23
SVD	Singular Value Decomposition	28
SSS	State of Self-Stress	27
IM	Inextensional Mechanism	27
MC	Monte Carlo	31
WCA	Weeks-Chandler-Anderson	53
ROC	Receiver Operating Characteristic	90
AUROC	Area Under Receiver Operating Characteristic curve	91
BLAST	Basic Local Alignment Search Tool	91
OUP	Ornstein-Uhlenbeck Process	45

List of Figures

List of Tables

Appendix A

Appendix

A.1 Semiempirical potential energy function

These potential energy function used in MD simulations describe in high detail the molecular interactions between the atoms in an assembly of molecules and have the form

$$
\begin{aligned}
U_{\text{ff}}(\boldsymbol{r}) \;=\; & \sum_{b} \frac{1}{2} K_b (b - b_0)^2 \\
& + \sum_{\theta} \frac{1}{2} K_\theta (\theta - \theta_0)^2 \\
& + \sum_{\zeta} \frac{1}{2} K_\zeta (\zeta - \zeta_0)^2 \\
& + \sum_{\phi} K_\phi (1 + \cos(n\phi - \delta)) \\
& + \sum_{\text{pairs: } ij} \frac{q_i q_j}{4\pi\epsilon_r \epsilon_0 r_{ij}} \\
& + \sum_{\text{pairs: } ij} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]
\end{aligned}
\tag{A.1}
$$

.

The terms can be separated into bonded and non bonded interactions. The first four terms in Eq. (A.1) represent the bonded interactions, summing over the bonds b, the angles θ, the improper dihedral angles ζ and the dihedral angles ϕ. The angle δ is the regular value of

the respective dihedral angle, r_{ij} is the distance between the two atoms i and j. The non-bonded interactions are taken into account with the last two terms in Eq. (A.1), where the sum is taken over all pairs of atoms i and j. They represent the Coulomb and the Lennard-Jones potential, where the latter approximates quantum mechanical effects. It consists of an attractive part, representing the London dispersion forces (van der Waals forces) and a repulsive term, accounting for the Pauli exclusion principle. ϵ_r is the dimensionless relative permittivity of the material in which the charges q_i are immersed, ϵ_0 is the electric constant, ε_{ij} is the depth of the Lennard-Jones potential well, σ_{ij} is the distance at which the inter-particle potential is zero. A nontrivial amount of both theoretical and experimental studies is devoted to deriving the constants in Eq. (A.1). The routines for NMA are for quite some time included in state of the art MD software [Van Der Spoel et al., 2005, Hess et al., 2008, Pronk et al., 2013, Abraham et al., 2015], often different potentials and solvent models are available.

A.2 Eigenvalue spectra

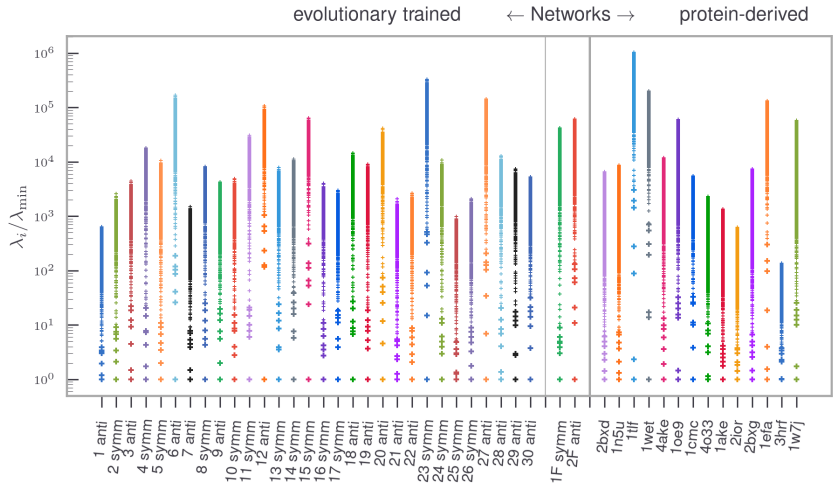


Figure A.1: The eigenvalues of the Hessian matrices are plotted on a logarithmic scale, normalized with the first non-zero eigenvalue. Spectra of all studied networks are determined, different networks are indicated with colors. Artificial networks are shown on the left, including those from [Flechsig, 2017] and protein-derived networks on the right. Note that there are no qualitative differences between these sets. A spectral gap between the first and subsequent eigenvalues can be observed for some networks, however, this is apparently not a necessary condition for allostery.

A.3 Structural data

A.3.1 Proteins from the PDB

Table A.1: Overview over the parameters of the proteins we analyzed in this work. r_c is the cutoff length, in Å. N is the total number of beads, N_T and N_S the number of beads composing the target and source pocket, respectively. N_s and N_p are the numbers of pocket candidates, subsampled pockets and original pockets, respectively.

Name	PDB	state	r_c	N	N_T	N_S	N_s	N_p	Reference
HSP90	2ior	holo	9	228	3	9	1397	33	[Shiau et al., 2006]
MetJ	1cmc	apo	10	104	3	10	640	23	[Rafferty et al., 1989]
PurR	1wet	holo	11	338	27	5	2776	48	[Schumacher et al., 1997]
HSA	1n5u	holo	9	583	16	14	3523	51	[Wardell et al., 2002]
HSA	2bxd	holo	9	578	16	8	4169	77	[Ghuman et al., 2005]
HSA	2bxg	holo	9	578	16	14	3168	68	[Ghuman et al., 2005]
ADK	4ake	apo	8	214	4	5	1390	27	[Müller et al., 1996]
ADK	1ake	holo	8	214	4	5	1145	29	[Müller and Schulz, 1992]
MyoV	1oe9	apo	9	730	3	31	5192	110	[Coureux et al., 2003]
MyoV	1w7j	holo	9	752	3	31	6065	116	[Coureux et al., 2004]
PGK1	4o33	holo	9	417	6	21	3029	53	[Chen et al., 2015]
PDK1	3hrf	holo	14	287	4	10	1388	24	[Hindie et al., 2009]
LacR	1efa	holo	10	328	10	14	2154	47	[Bell and Lewis, 2000]
LacR	1tlf	apo	10	296	10	14	1161	47	[Friedman et al., 1995]

A.4 Artificial allosteric networks

Table A.2: Overview over the parameters of the networks we analyzed in this work. r_c is the cutoff length, unitless. N is the total number of beads, N_T and N_S the number of beads composing the target and source pocket, respectively. N_{sp} and N_{st} are the numbers of pocket candidates, bead pairs and triplets, respectively.

ID	variant	r_c	N	N_T	N_S	N_{sp}	N_{st}	origin
1	anti	1.63	120	2	2	206	206	packed spheres
2	symm	6.0	95	3	3	376	376	trained pseudo protein
3	anti	6.0	94	3	3	295	295	trained pseudo protein
4	symm	6.0	89	3	3	224	224	trained pseudo protein
5	symm	6.0	142	3	3	201	201	trained pseudo protein
6	anti	6.0	142	3	3	309	309	trained pseudo protein
7	anti	6.0	180	3	3	530	530	trained pseudo protein
8	symm	6.0	93	3	3	461	461	trained pseudo protein
9	anti	6.0	154	3	3	315	315	trained pseudo protein
10	symm	6.0	100	3	3	540	540	trained pseudo protein
11	symm	6.0	99	3	3	401	401	trained pseudo protein
12	anti	6.0	94	3	3	294	294	trained pseudo protein
13	symm	6.0	93	3	3	535	535	trained pseudo protein
14	symm	6.0	147	3	2	495	495	trained pseudo protein
15	symm	6.0	94	3	3	330	330	trained pseudo protein
16	symm	6.0	99	3	3	332	332	trained pseudo protein
17	symm	6.0	154	3	3	465	465	trained pseudo protein
18	anti	6.0	89	3	3	288	288	trained pseudo protein
19	anti	6.0	93	3	3	461	461	trained pseudo protein
20	anti	6.0	95	3	3	317	317	trained pseudo protein
21	anti	6.0	99	3	3	400	400	trained pseudo protein
22	anti	6.0	100	3	3	480	480	trained pseudo protein
23	symm	6.0	135	3	3	571	571	trained pseudo protein
24	symm	6.0	88	3	3	187	187	trained pseudo protein
25	symm	6.0	87	3	3	244	244	trained pseudo protein
26	symm	6.0	169	3	3	386	386	trained pseudo protein
27	anti	6.0	147	3	2	427	427	trained pseudo protein
28	anti	6.0	99	3	3	353	353	trained pseudo protein
29	anti	6.0	87	3	3	234	234	trained pseudo protein
30	anti	6.0	169	3	3	382	382	trained pseudo protein
1F	anti	9.0	200	2	2	845	2204	[Flechsig, 2017]
2F	symm	9.0	200	2	2	892	2691	[Flechsig, 2017]

A.5 Energy rise during response

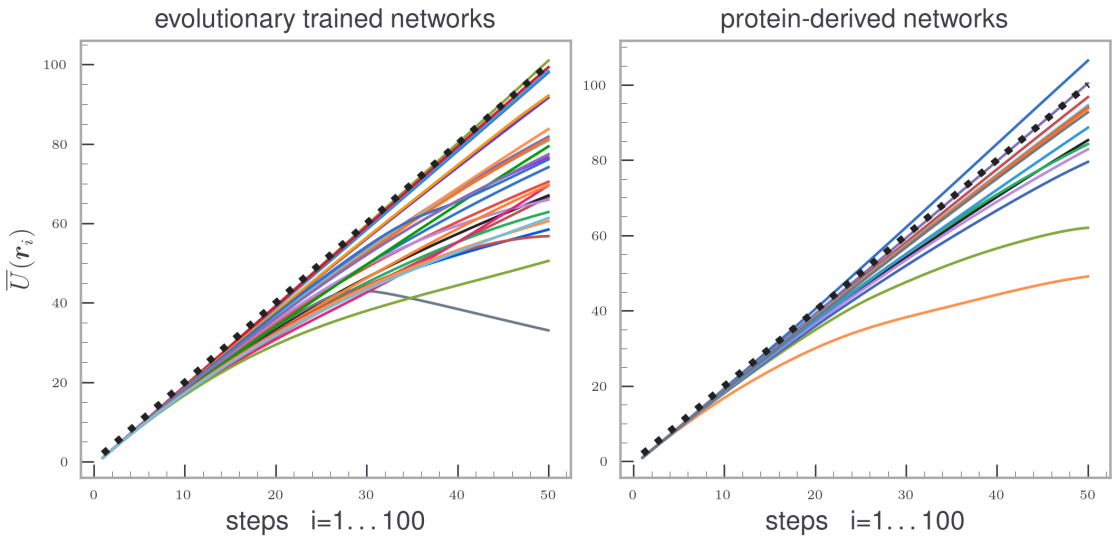
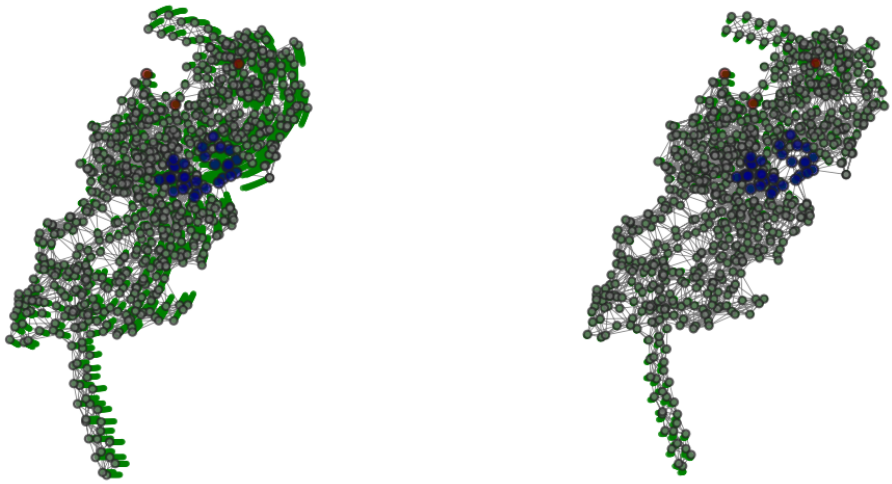


Figure A.2: Slope of the energy rise during response, normalized with the slope during the first step. Left for the artificial networks, right for the protein-derived networks. Deviations from a pure quadratic rise are visible.

A.6 Non-reciprocity in myosin V



(a) Perturbation at the allosteric site **(b)** Outer sphere

Figure A.3: Non-reciprocity in the response of myosin V [Coureux et al., 2004].

A.7 Comparing heatbath with allosteric effect

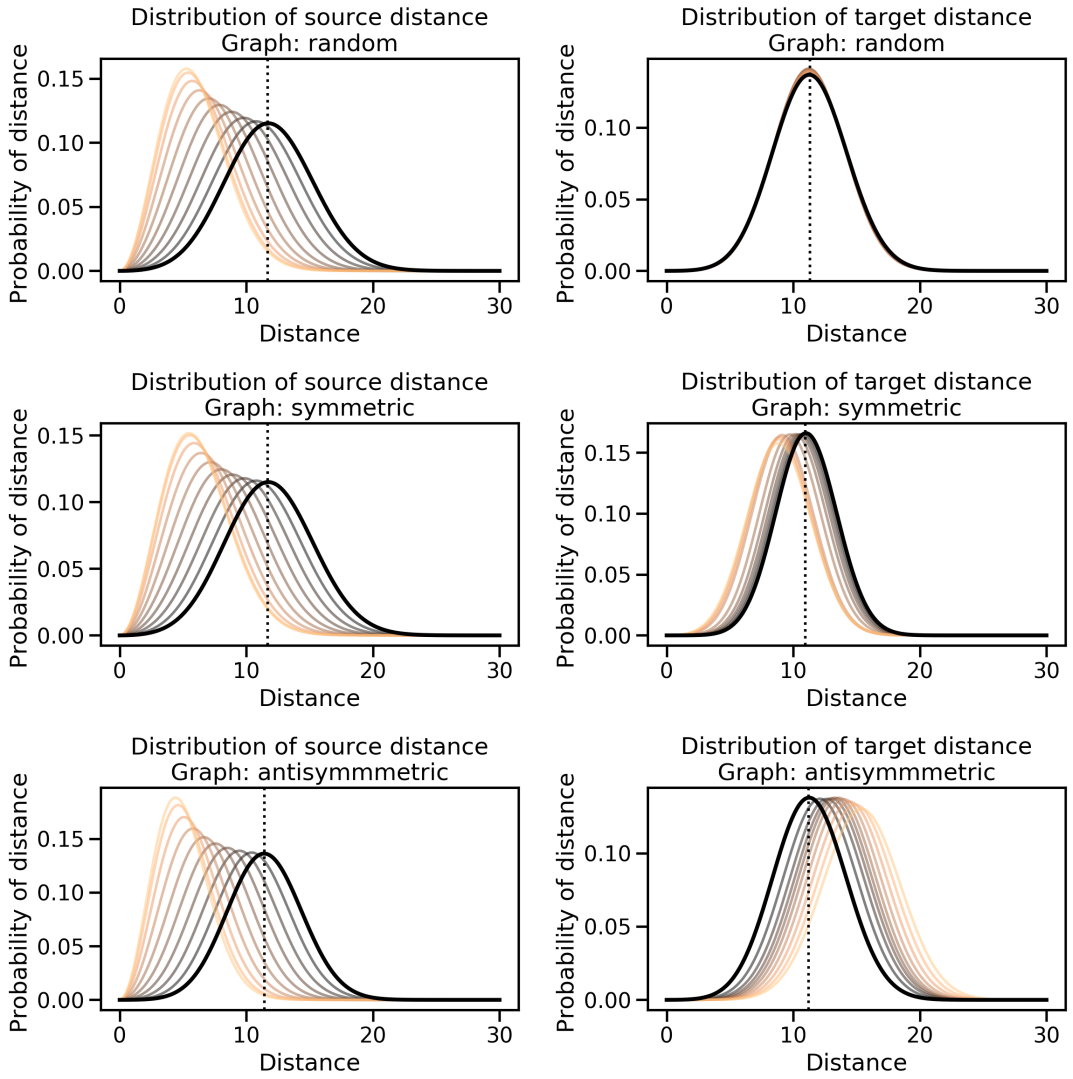


Figure A.4: To obtain distance fluctuations during the response, we evaluate the equilibrium fluctuations as derived in section 4.3 along the computed path.

A.8 Local spectra

A.8.1 Evolutionary trained networks

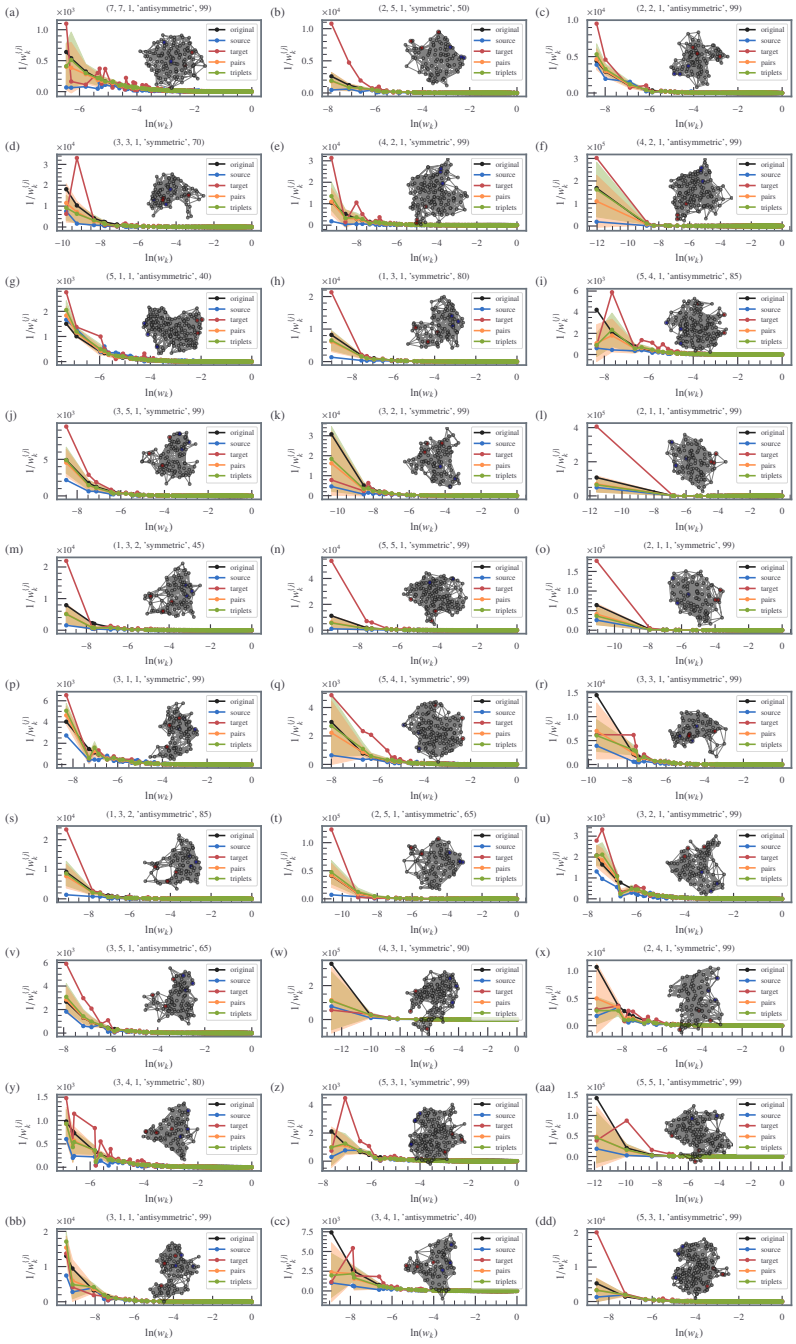


Figure A.5: Overview of local spectra in evolutionary trained networks

A.8.2 Protein-derived networks

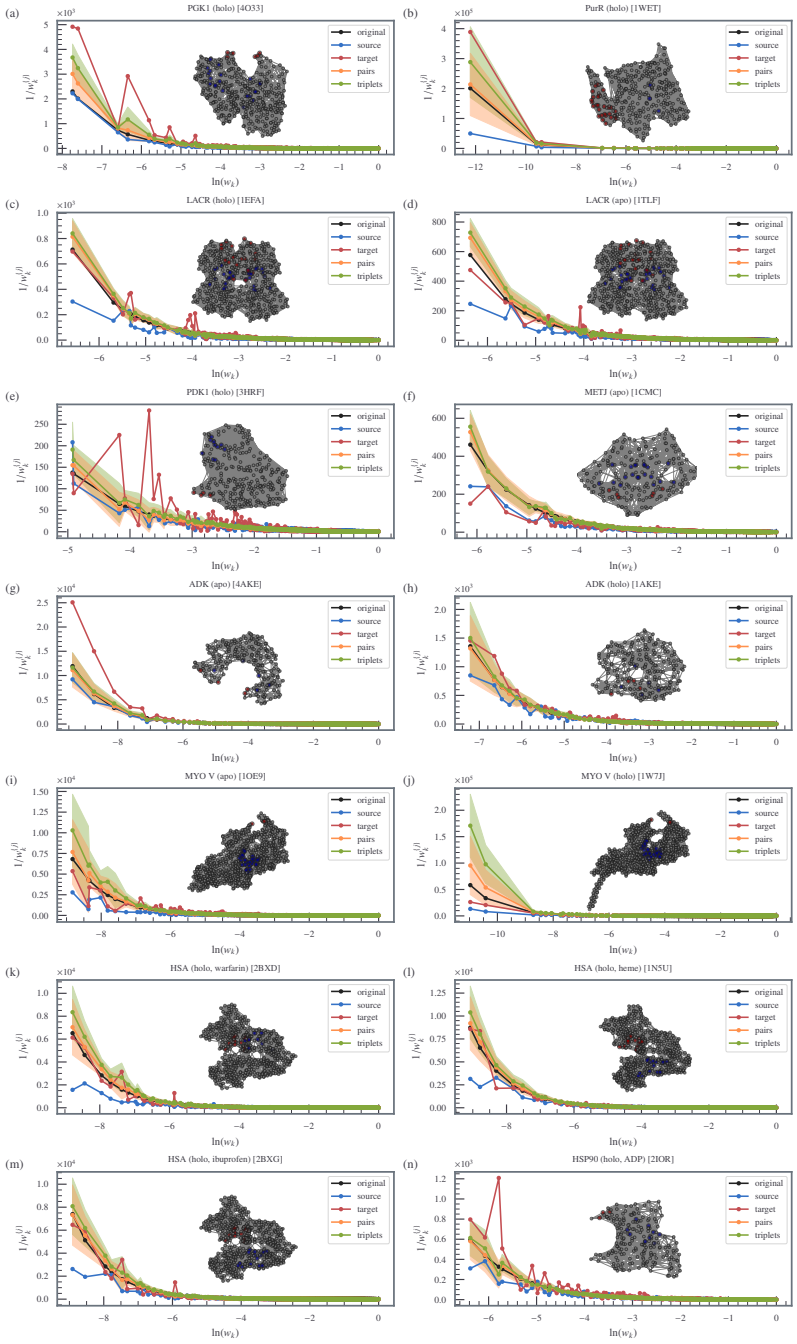


Figure A.6: Overview of local spectra in protein-derived networks

A.9 Pocket size distributions

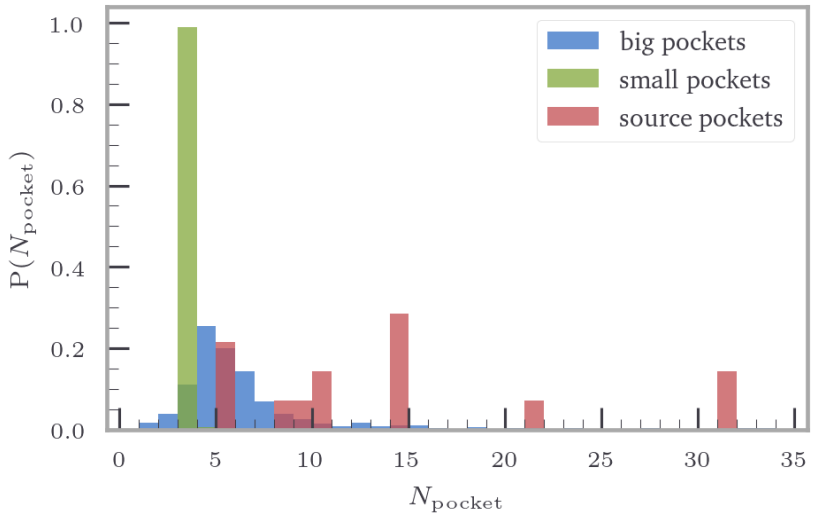


Figure A.7: Pocket size distributions for the proteins.

A.10 ROC curves for energy change during constraining

The ROC curve for the prediction based on energy input during the constraining step.

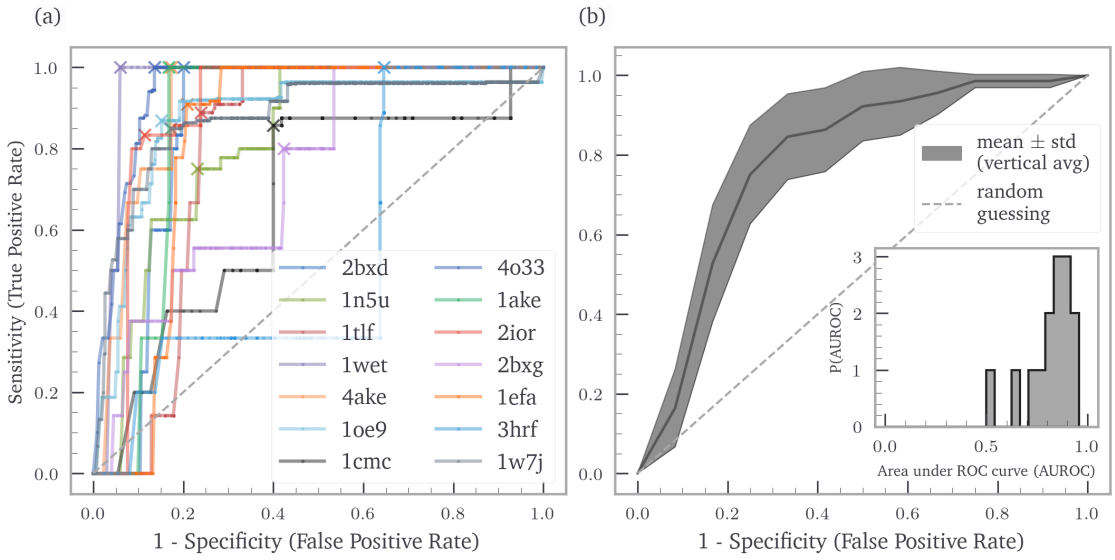


Figure A.8: For an explanation refer to Fig. 6.16. Results are similar on average but with a larger spread.

A.11 Smooth energy landscape

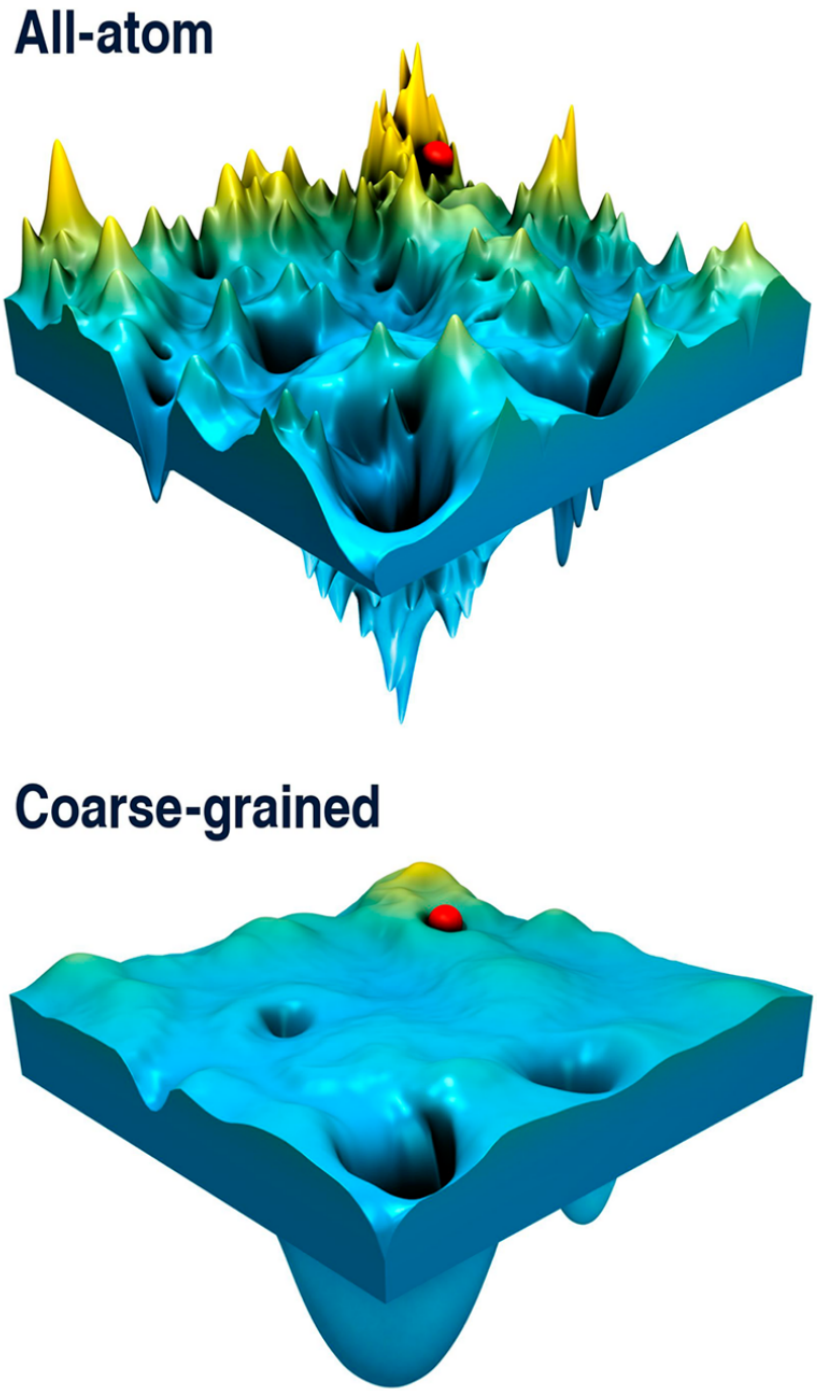


Figure A.9: Coarse-grained models smoothen the energy landscape, small barriers and local minima effectively disappear. Figure from [Kmiecik et al., 2016]

Appendix B

Software

After publication, the software accompanying this thesis will be available as an open-source python package at:

`https://gitlab.gwdg.de/mvossel/elastory`.

Bibliography

[Abraham et al., 2015] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25. doi:10.1016/j.softx.2015.06.001.

[Alberts et al., 2020] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2020). *Molecular Biology of the Cell (Sixth Edition) EBook Folder*. W. W. Norton, Incorporated. URL: https://books.google.de/books?id=uD-KzQEACAAJ.

[Amor et al., 2016] Amor, B. R. C., Schaub, M. T., Yaliraki, S. N., and Barahona, M. (2016). Prediction of allosteric sites and mediating interactions through bond-to-bond propensities. *Nature Communications*, 7(1):12477. URL: http://www.nature.com/articles/ncomms12477, doi:10.1038/ncomms12477.

[Anderson, 1958] Anderson, P. W. (1958). Absence of diffusion in certain random lattices. *Physical review*, 109(5):1492. doi:10.1142/9789812567154_0007.

[Anderson, 1978] Anderson, P. W. (1978). Local moments and localized states. *Science*, 201(4353):307–316. doi:10.1126/science.201.4353.307.

[Ansari, 1999] Ansari, A. (1999). Langevin modes analysis of myoglobin. *The Journal of chemical physics*, 110(3):1774–1780. doi:10.1063/1.477885.

[Ansari et al., 1992] Ansari, A., Jones, C. M., Henry, E. R., Hofrichter, J., and Eaton, W. A. (1992). The role of solvent viscosity in the dynamics of protein conformational changes. *Science*, 256(5065):1796–1798. doi:10.1126/science.1615323.

[Ascenzi and Fasano, 2010] Ascenzi, P. and Fasano, M. (2010). Allostery in a monomeric protein: The case of human serum albumin. *Biophysical Chemistry*, 148:7. doi:10.1016/j.bpc.2010.03.001.

[Asthana et al., 2007] Asthana, S., Roytberg, M., Stamatoyannopoulos, J., and Sunyaev, S. (2007). Analysis of sequence conservation at nucleotide resolution. *PLoS computational biology*, 3(12):e254. doi:10.1371/journal.pcbi.0030254.

[Atilgan et al., 2007] Atilgan, A. R., Turgut, D., and Atilgan, C. (2007). Screened nonbonded interactions in native proteins manipulate optimal paths for robust residue communication. *Biophysical journal*, 92(9):3052–3062. URL: http://www.cell.com/article/S0006349507711141/pdf, doi:10.1529/biophysj.106.099440.

[Atilgan and Atilgan, 2009] Atilgan, C. and Atilgan, A. R. (2009). Perturbation-Response Scanning Reveals Ligand Entry-Exit Mechanisms of Ferric Binding Protein. *PLoS Computational Biology*, 5(10):e1000544. URL: https://dx.plos.org/10.1371/journal.pcbi.1000544, doi:10.1371/journal.pcbi.1000544.

[Bahar et al., 1998] Bahar, I., Atilgan, A. R., Demirel, M. C., and Erman, B. (1998). Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Physical Review Letters*, 80(12):2733. doi:10.1103/physrevlett.80.2733.

[Bahar and Jernigan, 1997] Bahar, I. and Jernigan, R. L. (1997). Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation. *Journal of molecular biology*, 266(1):195–214.

[Bahar et al., 2010] 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. URL: https://europepmc.org/articles/pmc2938190?pdf=render, doi:10.1146/annurev.biophys.093008.131258.

[Bakan et al., 2011] Bakan, A., Meireles, L. M., and Bahar, I. (2011). ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27(11):1575–1577. URL: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr168, doi:10.1093/bioinformatics/btr168.

[Baranau and Tallarek, 2017] Baranau, V. and Tallarek, U. (2017). Another resolution of the configurational entropy paradox as applied to hard spheres. *The Journal of Chemical Physics*, 147(22):224503. URL: http://arxiv.org/pdf/1706.09671, doi:10.1063/1.4999483.

[Bartlett et al., 2002] Bartlett, G. J., Porter, C. T., Borkakoti, N., and Thornton, J. M. (2002). Analysis of catalytic residues in enzyme active sites. *Journal of molecular biology*, 324(1):105–121. doi:10.1016/s0022-2836(02)01036-7.

[Bauer et al., 2019] Bauer, J. A., Pavlović, J., and Bauerová-Hlinková, V. (2019). Normal mode analysis as a routine part of a structural investigation. *Molecules*, 24(18):3293. URL: https://www.mdpi.com/1420-3049/24/18/3293/pdf, doi:10.3390/molecules24183293.

[Bell and Lewis, 2000] Bell, C. E. and Lewis, M. (2000). A closer view of the conformation of the Lac repressor bound to operator. *Nature structural biology*, 7(3):209–214.

[Ben-Avraham, 1993] Ben-Avraham, D. (1993). Vibrational normal-mode spectrum of globular proteins. *Physical Review B*, 47(21):14559. doi:10.1103/physrevb.47.14559.

[Bishop, 1984] Bishop, M. (1984). WCA perturbation theory for one-dimensional Lennard–Jones fluids. *American Journal of Physics*, 52(2):158–161. doi:10.1119/1.13728.

[Blázquez et al., 1996] Blázquez, A., Mantič, V., París, F., and Cañas, J. (1996). On the Removal of Rigid Body Motions in the Solution of Elastostatic Problems by Direct Bem. *International Journal for Numerical Methods in Engineering*, 39(23):4021–4038. doi:10.1002/(SICI)1097-0207(19961215)39:23<4021::AID-NME36>3.0.CO;2-Q.

[Bock et al., 2019] Bock, L. V., Caliskan, N., Korniy, N., Peske, F., Rodnina, M. V., and Grubmüller, H. (2019). Thermodynamic control of- 1 programmed ribosomal frameshifting. *Nature communications*, 10(1):1–11. URL: https://www.nature.com/articles/s41467-019-12648-x.pdf, doi:10.1038/s41467-019-12648-x.

[Boell et al., 1954] Boell, E. J. et al. (1954). Dynamics of growth processes. *Dynamics of growth processes.*, 36:166. URL: https://www.bio-conferences.org/10.1051/bioconf/20202505006/pdf, doi:10.2307/1931453.

[Bohr et al., 1904] Bohr, C., Hasselbalch, K., and Krogh, A. (1904). Ueber einen in biologischer Beziehung wichtigen Einfluss, den die Kohlensäurespannung des Blutes auf dessen Sauerstoffbindung übt 1. *Skandinavisches Archiv Für Physiologie*, 16(2):402–412.

[Brooks and Karplus, 1983] Brooks, B. and Karplus, M. (1983). Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences*, 80(21):6571–6575. URL: http://www.pnas.org/content/80/21/6571.full.pdf, doi:10.1073/pnas.80.21.6571.

[Brooks et al., 1983] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. a., and Karplus, M. (1983). CHARMM: a program for macromolecular en-

ergy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217. doi:10.1002/jcc.540040211.

[Brown, 1828] Brown, R. (1828). *A brief account of microscopical observations made in the months of June, July, and August, 1827*. A. and C. Black.

[Brown and Case, 2006] Brown, R. A. and Case, D. A. (2006). Second derivatives in generalized Born theory. *Journal of Computational Chemistry*, 27(14):1662–1675. doi:10.1002/jcc.20479.

[Bu and Callaway, 2011] Bu, Z. and Callaway, D. J. (2011). Proteins move! Protein dynamics and long-range allostery in cell signaling. *Advances in protein chemistry and structural biology*, 83:163–221. doi:10.1016/b978-0-12-381262-9.00005-7.

[Buhrow et al., 2012] Buhrow, L., Ferguson-Miller, S., and Kuhn, L. A. (2012). From static structure to living protein: computational analysis of cytochrome c oxidase main-chain flexibility. *Biophysical journal*, 102(9):2158–2166. URL: http://www.cell.com/article/S0006349511047424/pdf, doi:10.1016/j.bpj.2011.11.3394.

[Calladine, 1978] Calladine, C. (1978). Buckminster Fuller's "Tensegrity" structures and Clerk Maxwell's rules for the construction of stiff frames. *International Journal of Solids and Structures*, 14(2):161–172. URL: https://linkinghub.elsevier.com/retrieve/pii/0020768378900525, doi:10.1016/0020-7683(78)90052-5.

[Chandler et al., 1983] Chandler, D., Weeks, J. D., and Andersen, H. C. (1983). Van der Waals picture of liquids, solids, and phase transformations. *Science*, 220(4599):787–794. doi:10.1126/science.220.4599.787.

[Changeux, 1966] Changeux, J.-P. (1966). Responses of acetylcholinesterase from Torpedo marmorata to salts and curarizing drugs. *Molecular Pharmacology*, 2(5):369–392.

[Changeux, 2011] Changeux, J.-P. (2011). 50th anniversary of the word "allosteric". *Protein Science*, 20(7):1119–1124. URL: https://europepmc.org/articles/pmc3149185?pdf=render, doi:10.1002/pro.658.

[Changeux, 2012] Changeux, J.-P. (2012). Allostery and the Monod-Wyman-Changeux model after 50 years. *Annual review of biophysics*, 41:103–133. doi:10.1146/annurev-biophys-050511-102222.

[Changeux, 2013] Changeux, J.-P. (2013). 50 years of allosteric interactions: the twists and turns of the models. *Nature Reviews Molecular Cell Biology*, 14(12):819–829. URL: http://www.nature.com/articles/nrm3695, doi:10.1038/nrm3695.

[Chatzigoulas and Cournia, 2021] Chatzigoulas, A. and Cournia, Z. (2021). Rational design of allosteric modulators: Challenges and successes. *WIREs Computational Molecular Science*, 11(6). URL: https://onlinelibrary.wiley.com/doi/10.1002/wcms.1529, doi:10.1002/wcms.1529.

[Chen et al., 2015] Chen, X., Zhao, C., Li, X., Wang, T., Li, Y., Cao, C., Ding, Y., Dong, M., Finci, L., Wang, J.-h., et al. (2015). Terazosin activates Pgk1 and Hsp90 to promote stress resistance. *Nature chemical biology*, 11(1):19–25. URL: https://europepmc.org/articles/pmc4412158?pdf=render, doi:10.1038/nchembio.1657.

[Clapeyron and Lamé, 1831] Clapeyron, B. and Lamé, G. (1831). Mémoire sur l'équilibre intérieur des corps solides homogènes. 1831:145–169. URL: https://zenodo.org/record/1577754/files/article.pdf, doi:10.1515/crll.1831.7.145.

[Connolly, 1983] Connolly, M. L. (1983). Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713. doi:10.1126/science.6879170.

[Cooper and Dryden, 1984] Cooper, A. and Dryden, D. (1984). Allostery without conformational change. *European Biophysics Journal*, 11(2):103–109. doi:10.1007/bf00276625.

[Coulais et al., 2017] Coulais, C., Sounas, D., and Alu, A. (2017). Static non-reciprocity in mechanical metamaterials. *Nature*, 542(7642):461–464. URL: http://arxiv.org/pdf/1704.03305, doi:10.1038/nature21044.

[Coureux et al., 2004] Coureux, P.-D., Sweeney, H. L., and Houdusse, A. (2004). Three myosin V structures delineate essential features of chemo-mechanical transduction. *The EMBO journal*, 23(23):4527–4537. URL: https://europepmc.org/articles/pmc533045?pdf=render, doi:10.1038/sj.emboj.7600458.

[Coureux et al., 2003] Coureux, P.-D., Wells, A. L., Ménétrey, J., Yengo, C. M., Morris, C. A., Sweeney, H. L., and Houdusse, A. (2003). A structural state of the myosin V motor without bound nucleotide. *Nature*, 425(6956):419–423. doi:10.1038/nature01927.

[Crapo, 1979] Crapo, H. (1979). Structural rigidity. *Structural topology, 1979, núm. 1*. Publisher: Université du Québec à Montréal.

[Cuendet et al., 2016] Cuendet, M. A., Weinstein, H., and LeVine, M. V. (2016). The allostery landscape: Quantifying thermodynamic couplings in biomolecular systems. *Journal of chemical theory and computation*, 12(12):5758–5767. URL: http://www.cell.com/article/S0006349516329526/pdf, doi:10.1021/acs.jctc.6b00841.

[Cui and Bahar, 2005] Cui, Q. and Bahar, I. (2005). *Normal mode analysis: theory and applications to biological and chemical systems*. CRC press.

[Cui and Karplus, 2008] Cui, Q. and Karplus, M. (2008). Allostery and cooperativity revisited. *Protein science*, 17(8):1295–1307. URL: https://europepmc.org/articles/pmc2492820?pdf=render, doi:10.1110/ps.03259908.

[Daily and Gray, 2007] Daily, M. D. and Gray, J. J. (2007). Local motions in a benchmark of allosteric proteins. *Proteins: Structure, Function, and Bioinformatics*, 67(2):385–399. URL: https://onlinelibrary.wiley.com/doi/10.1002/prot.21300, doi:10.1002/prot.21300.

[De Nittis and Lein, 2017] De Nittis, G. and Lein, M. (2017). *Linear response theory: an analytic-algebraic approach*, volume 21. Springer.

[Deb et al., 2009] Deb, D., Vishveshwara, S., and Vishveshwara, S. (2009). Understanding protein structure from a percolation perspective. *Biophysical journal*, 97(6):1787–1794. URL: http://www.cell.com/article/S0006349509012387/pdf, doi:10.1016/j.bpj.2009.07.016.

[Debye, 1912] Debye, P. (1912). Zur theorie der spezifischen wärmen. *Annalen der Physik*, 344(14):789–839. URL: https://zenodo.org/record/1424256/files/article.pdf, doi:10.1002/andp.19123441404.

[Deiters and Randzio, 1995] Deiters, U. and Randzio, S. (1995). The equation of state for molecules with shifted Lennard-Jones pair potentials. *Fluid phase equilibria*, 103(2):199–212. doi:10.1016/0378-3812(94)02577-n.

[del Sol et al., 2006] del Sol, A., Fujihashi, H., Amoros, D., and Nussinov, R. (2006). Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Molecular Systems Biology*, 2(1). URL: https://onlinelibrary.wiley.com/doi/abs/10.1038/msb4100063, doi:10.1038/msb4100063.

[Delanuay, 1934] Delanuay, B. (1934). Sur la sphere vide. A la memoire de Georges Voronoi. *Proceedings of the USSR Academy of Sciences*, (6):793–800.

[Detweiler et al., 2007] Detweiler, C., Vona, M., Yoon, Y., Yun, S.-k., and Rus, D. (2007). Self-assembling mobile linkages. *IEEE robotics and automation magazine*, 14(4):45–55. doi:10.1109/mra.2007.4383467.

[Diamond, 1990] Diamond, R. (1990). On the use of normal modes in thermal parameter refinement: theory and application to the bovine pancreatic trypsin inhibitor.

Acta Crystallographica Section A: Foundations of Crystallography, 46(6):425–435. doi:10.1107/s0108767390002082.

[Dixit and Verkhivker, 2011] Dixit, A. and Verkhivker, G. M. (2011). Computational modeling of allosteric communication reveals organizing principles of mutation-induced signaling in ABL and EGFR kinases. *PLoS computational biology*, 7(10):e1002179. doi:10.1371/journal.pcbi.1002179.

[Dokholyan, 2016] Dokholyan, N. V. (2016). Controlling Allosteric Networks in Proteins. *Chemical reviews*, 116(11):6463–6487. doi:10.1021/acs.chemrev.5b00544.

[Düring et al., 2013] Düring, G., Lerner, E., and Wyart, M. (2013). Phonon gap and localization lengths in floppy materials. *Soft Matter*, 9(1):146–154. URL: http://arxiv.org/pdf/1204.3542, doi:10.1039/c2sm25878a.

[Edelsbrunner, 1993] Edelsbrunner, H. (1993). The union of balls and its dual shape. In *Proceedings of the ninth annual symposium on Computational geometry*, volume 13, pages 218–231. Springer Science and Business Media LLC. URL: https://link.springer.com/content/pdf/10.1007%2FBF02574053.pdf, doi:10.1007/bf02574053.

[Edelsbrunner and Mücke, 1994] Edelsbrunner, H. and Mücke, E. P. (1994). Three-dimensional alpha shapes. *ACM Transactions on Graphics (TOG)*, 13(1):43–72. URL: http://arxiv.org/pdf/math/9410208v1.pdf, doi:10.1145/147130.147153.

[Eigen, 1968] Eigen, M. (1968). New looks and outlooks on physical enzymology. *Quarterly reviews of biophysics*, 1(1):3–33. doi:10.1017/s0033583500000445.

[Einstein, 1905] Einstein, A. (1905). *Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*. J. Barth.

[Engelke, 2021] Engelke, L. (2021). Fused Spheres. https://gitlab.com/engelke/fusedspheres. URL: https://gitlab.com/engelke/fusedspheres.

[Eyal and Bahar, 2008] Eyal, E. and Bahar, I. (2008). Toward a Molecular Understanding of the Anisotropic Response of Proteins to External Forces: Insights from Elastic Network Models. *Biophysical Journal*, 94(9):3424–3435. URL: http://www.cell.com/article/S0006349508704223/pdf, doi:10.1529/biophysj.107.120733.

[Eyal et al., 2006] Eyal, E., Yang, L.-W., and Bahar, I. (2006). Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, 22(21):2619–2627.

URL: https://academic.oup.com/bioinformatics/article-pdf/22/21/2619/697193/btl448.pdf, doi:10.1093/bioinformatics/btl448.

[Fawcett, 2006] Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874. URL: http://www.decom.ufop.br/menotti/rp112/slides/13-2006-AnIntroductionToROCAnalysis-PRLv27pp861-874.pdf, doi:10.1016/j.patrec.2005.10.010.

[Fealey and Hinderliter, 2013] Fealey, M. E. and Hinderliter, A. (2013). Allostery and instability in the functional plasticity of synaptotagmin I. *Communicative and Integrative Biology*, 6(2):e46748. URL: https://www.tandfonline.com/doi/pdf/10.4161/cib.22830?needAccess=true, doi:10.4161/cib.22830.

[Fenimore et al., 2004] Fenimore, P. W., Frauenfelder, H., McMahon, B., and Young, R. (2004). Bulk-solvent and hydration-shell fluctuations, similar to α-and β-fluctuations in glasses, control protein motions and functions. *Proceedings of the National Academy of Sciences*, 101(40):14408–14413. doi:10.1073/pnas.0405573101.

[Fenimore et al., 2002] Fenimore, P. W., Frauenfelder, H., McMahon, B. H., and Parak, F. G. (2002). Slaving: solvent fluctuations dominate protein dynamics and functions. *Proceedings of the National Academy of Sciences*, 99(25):16047–16051. URL: http://www.pnas.org/content/99/25/16047.full.pdf, doi:10.1073/pnas.212637899.

[Fenton, 2008] Fenton, A. W. (2008). Allostery: an illustrated definition for the 'second secret of life'. *Trends in Biochemical Sciences*, 33(9):420–425. URL: https://linkinghub.elsevier.com/retrieve/pii/S0968000408001643, doi:10.1016/j.tibs.2008.05.009.

[Flechsig, 2017] Flechsig, H. (2017). Design of Elastic Networks with Evolutionary Optimized Long-Range Communication as Mechanical Models of Allosteric Proteins. *Biophysical Journal*, 113(3):558–571. URL: http://www.cell.com/article/S0006349517306938/pdf, doi:10.1016/j.bpj.2017.06.043.

[Flory, 1944] Flory, P. J. (1944). Network Structure and the Elastic Properties of Vulcanized Rubber. *Chemical Reviews*, 35(1):51–75. URL: https://pubs.acs.org/doi/abs/10.1021/cr60110a002, doi:10.1021/cr60110a002.

[Flory, 1976] Flory, P. J. (1976). Statistical Thermodynamics of Random Networks. *Proc. R. Soc. Lond. A*, 351(1666):351–380. URL: https://royalsocietypublishing.org/doi/10.1098/rspa.1976.0146, doi:10.1098/rspa.1976.0146.

[Fogolari et al., 2012] Fogolari, F., Corazza, A., Viglino, P., and Esposito, G. (2012). Fast structure similarity searches among protein models: efficient clustering of protein fragments. *Algorithms for Molecular Biology*, 7(1):1–10. doi:10.1186/1748-7188-7-16.

[Formaneck et al., 2006] Formaneck, M. S., Ma, L., and Cui, Q. (2006). Reconciling the "old" and "new" views of protein allostery: a molecular simulation study of chemotaxis Y protein (CheY). *Proteins: Structure, Function, and Bioinformatics*, 63(4):846–867. doi:10.1002/prot.20893.

[Frauenfelder et al., 1991] Frauenfelder, H., Sligar, S. G., and Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603. doi:10.1126/science.1749933.

[Friedman et al., 1995] Friedman, A. M., Fischmann, T. O., and Steitz, T. A. (1995). Crystal structure of lac repressor core tetramer and its implications for DNA looping. *Science*, 268(5218):1721–1727. doi:10.1126/science.7792597.

[Garcia et al., 2012] Garcia, H. G., Sanchez, A., Boedicker, J. Q., Osborne, M., Gelles, J., Kondev, J., and Phillips, R. (2012). Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell reports*, 2(1):150–161. doi:10.1016/j.celrep.2012.06.004.

[Gardiner et al., 1985] Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 3. springer Berlin. doi:10.1007/978-3-662-02377-8.

[Gerstein and Krebs, 1998] Gerstein, M. and Krebs, W. (1998). A database of macromolecular motions. *Nucleic acids research*, 26(18):4280–4290. URL: https://academic.oup.com/nar/article-pdf/26/18/4280/9482770/26-18-4280.pdf, doi:10.1093/nar/26.18.4280.

[Ghuman et al., 2005] Ghuman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A., Otagiri, M., and Curry, S. (2005). Structural basis of the drug-binding specificity of human serum albumin. *Journal of molecular biology*, 353(1):38–52. doi:10.1016/j.jmb.2005.07.075.

[Gilson et al., 1997] Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophysical journal*, 72(3):1047–1069. doi:10.1016/s0006-3495(97)78756-3.

[Gniewek et al., 2012] Gniewek, P., Kolinski, A., Jernigan, R. L., and Kloczkowski, A. (2012). Elastic network normal modes provide a basis for protein structure refine-

ment. *The Journal of Chemical Physics*, 136(19):05B616. URL: https://europepmc.org/articles/pmc3365912?pdf=render, doi:10.1063/1.4710986.

[Go et al., 1983] 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*, 80(12):3696–3700. doi:10.1073/pnas.80.12.3696.

[Goldstein et al., 2002] Goldstein, H., Poole, C., and Safko, J. (2002). Classical mechanics. *Classical mechanics (3rd ed.) by H. Goldstein*, pages 18–58. doi:10.1142/9789814623957_0002.

[Goodrich et al., 2015] Goodrich, C. P., Liu, A. J., and Nagel, S. R. (2015). The principle of independent bond-level response: Tuning by pruning to exploit disorder for global behavior. *Physical review letters*, 114(22):225501. URL: https://link.aps.org/accepted/10.1103/PhysRevLett.114.225501, doi:10.1103/physrevlett.114.225501.

[Greener et al., 2017] Greener, J. G., Filippis, I., and Sternberg, M. J. (2017). Predicting Protein Dynamics and Allostery Using Multi-Protein Atomic Distance Constraints. *Structure*, 25(3):546–558. URL: http://www.cell.com/article/S0969212617300084/pdf, doi:10.1016/j.str.2017.01.008.

[Guarnera and Berezovsky, 2016] Guarnera, E. and Berezovsky, I. N. (2016). Allosteric sites: remote control in regulation of protein activity. *Current opinion in structural biology*, 37:1–8. doi:10.1016/j.sbi.2015.10.004.

[Guarnera and Berezovsky, 2019] Guarnera, E. and Berezovsky, I. N. (2019). Toward comprehensive allosteric control over protein activity. *Structure*, 27(5):866–878. URL: http://www.cell.com/article/S0969212619300140/pdf, doi:10.1016/j.str.2019.01.014.

[Gunasekaran et al., 2004] Gunasekaran, K., Ma, B., and Nussinov, R. (2004). Is allostery an intrinsic property of all dynamic proteins? *Proteins: Structure, Function, and Bioinformatics*, 57(3):433–443. doi:10.1002/prot.20232.

[Guo and Zhou, 2016] Guo, C. and Zhou, H.-X. (2016). Unidirectional allostery in the regulatory subunit RI-alpha facilitates efficient deactivation of protein kinase A. *Proceedings of the National Academy of Sciences*, 113(44):E6776–E6785.

[Güntert, 2009] Güntert, P. (2009). Automated structure determination from NMR spectra. *European Biophysics Journal*, 38(2):129–143. doi:10.1007/s00249-008-0367-z.

[Haliloglu et al., 1997] Haliloglu, T., Bahar, I., and Erman, B. (1997). Gaussian Dynamics of Folded Proteins. *Physical Review Letters*, 79(16):3090–3093. doi:10.1103/PhysRevLett.79.3090.

[Halle, 2002] Halle, B. (2002). Flexibility and packing in proteins. *Proceedings of the National Academy of Sciences*, 99(3):1274–1279. URL: https://europepmc.org/articles/pmc122180?pdf=render, doi:10.1073/pnas.032522499.

[Hayward and Groot, 2008] Hayward, S. and Groot, B. L. d. (2008). Normal modes and essential dynamics. In *Molecular Modeling of Proteins*, pages 89–106. Springer. doi:10.1007/978-1-59745-177-2_5.

[Hayward et al., 1993] Hayward, S., Kitao, A., Hirata, F., and Gō, N. (1993). Effect of solvent on collective motions in globular protein. *Journal of molecular biology*, 234(4):1207–1217. doi:10.1006/jmbi.1993.1671.

[Hess et al., 2008] Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of chemical theory and computation*, 4(3):435–447. URL: https://pure.mpg.de/pubman/item/item_588952_3/component/file_588951/412029.pdf, doi:10.1021/ct700301q.

[Hill, 1910] Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *j. physiol.*, 40:4–7.

[Hilser et al., 2012] Hilser, V. J., Wrabl, J. O., and Motlagh, H. N. (2012). Structural and energetic basis of allostery. *Annual review of biophysics*, 41:585–609. URL: https://europepmc.org/articles/pmc3935618?pdf=render, doi:10.1146/annurev-biophys-050511-102319.

[Hindie et al., 2009] Hindie, V., Stroba, A., Zhang, H., Lopez-Garcia, L. A., Idrissova, L., Zeuzem, S., Hirschberg, D., Schaeffer, F., Jørgensen, T. J., Engel, M., et al. (2009). Structure and allosteric effects of low-molecular-weight activators on the protein kinase PDK1. *Nature chemical biology*, 5(10):758–764. doi:10.1038/nchembio.208.

[Hinsen, 1998] Hinsen, K. (1998). Analysis of domain motions by approximate normal mode calculations. *Proteins: Structure, Function, and Bioinformatics*, 33(3):417–429. doi:10.1002/(sici)1097-0134(19981115)33:3<417::aid-prot10>3.0.co;2-8.

[Hinsen et al., 2000] Hinsen, K., Petrescu, A.-J., Dellerue, S., Bellissent-Funel, M.-C., and Kneller, G. R. (2000). Harmonicity in slow protein dynamics. *Chemical Physics*, 261(1-2):25–37. doi:10.1016/s0301-0104(00)00222-6.

[Hong et al., 2012] Hong, L., Cheng, X., Glass, D. C., and Smith, J. C. (2012). Surface hydration amplifies single-well protein atom diffusion propagating into the macromolecular core. *Physical review letters*, 108(23):238102. URL: https://link.aps.org/accepted/10.1103/PhysRevLett.108.238102, doi:10.1103/physrevlett.108.238102.

[Hosokawa et al., 2021] Hosokawa, N., Kuragano, M., Yoshino, A., Shibata, K., Uyeda, T. Q., and Tokuraku, K. (2021). Unidirectional cooperative binding of fimbrin actin-binding domain 2 to actin filament. *Biochemical and Biophysical Research Communications*, 552:59–65. doi:10.1016/j.bbrc.2021.02.139.

[Hospital et al., 2015] Hospital, A., Goni, J. R., Orozco, M., and Gelpi, J. L. (2015). Molecular dynamics simulations: advances and applications. *Advances and applications in bioinformatics and chemistry: AABC*, 8:37. URL: https://www.dovepress.com/getfile.php?fileID=28066, doi:10.2147/aabc.s70333.

[Huang et al., 2013] Huang, W., Lu, S., Huang, Z., Liu, X., Mou, L., Luo, Y., Zhao, Y., Liu, Y., Chen, Z., Hou, T., et al. (2013). Allosite: a method for predicting allosteric sites. *Bioinformatics*, 29(18):2357–2359. URL: https://academic.oup.com/bioinformatics/article-pdf/29/18/2357/17128673/btt399.pdf, doi:10.1093/bioinformatics/btt399.

[Hüfner, 1890] Hüfner, G. (1890). Über das Gesetz der Dissociation des Oxyhaemoglobins und Über einige daran sich knüpfende wichtige Fragen aus der Biologie. *Arch. Anat. Physiol*, 1:1–27.

[Ikeguchi et al., 2005] Ikeguchi, M., Ueno, J., Sato, M., and Kidera, A. (2005). Protein Structural Change Upon Ligand Binding: Linear Response Theory. *Physical Review Letters*, 94(7):078102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.94.078102, doi:10.1103/PhysRevLett.94.078102.

[Kabsch, 1976] Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):922–923. doi:10.1107/s0567739476001873.

[Kar et al., 2010] Kar, G., Keskin, O., Gursoy, A., and Nussinov, R. (2010). Allostery and population shift in drug discovery. *Current opinion in pharmacology*, 10(6):715–722. URL: https://europepmc.org/articles/pmc7316380?pdf=render, doi:10.1016/j.coph.2010.09.002.

[Kaya et al., 2013] Kaya, C., Armutlulu, A., Ekesan, S., and Haliloglu, T. (2013). MCPath: Monte Carlo path generation approach to predict likely allosteric pathways and functional residues. *Nucleic acids research*, 41(W1):W249–W255. URL: https://academic.oup.com/nar/article-pdf/41/W1/W249/16941844/gkt284.pdf, doi:10.1093/nar/gkt284.

[Kidera and Gō, 1992] Kidera, A. and Gō, N. (1992). Normal mode refinement: crystallographic refinement of protein dynamic structure: I. Theory and test by simulated diffraction data. *Journal of molecular biology*, 225(2):457–475.

[Kim et al., 2022] Kim, J. Z., Lu, Z., Blevins, A. S., and Bassett, D. S. (2022). Nonlinear Dynamics and Chaos in Conformational Changes of Mechanical Metamaterials. *Physical Review X*, 12(1):011042. URL: http://link.aps.org/pdf/10.1103/PhysRevX.12.011042, doi:10.1103/physrevx.12.011042.

[Kim et al., 2019] Kim, J. Z., Lu, Z., Strogatz, S. H., and Bassett, D. S. (2019). Conformational control of mechanical networks. *Nature Physics*, 15:1. URL: http://arxiv.org/pdf/1804.00173, doi:10.1038/s41567-019-0475-y.

[Kim et al., 2013] Kim, S., Broströmer, E., Xing, D., Jin, J., Chong, S., Ge, H., Wang, S., Gu, C., Yang, L., Gao, Y. Q., et al. (2013). Probing allostery through DNA. *Science*, 339(6121):816–819. URL: https://europepmc.org/articles/pmc3586787?pdf=render, doi:10.1126/science.1229223.

[Kimura et al., 1968] Kimura, M. et al. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626. doi:10.1038/217624a0.

[King and Jukes, 1969] King, J. L. and Jukes, T. H. (1969). Non-Darwinian Evolution: Most evolutionary change in proteins may be due to neutral mutations and genetic drift. *Science*, 164(3881):788–798.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. URL: http://arxiv.org/abs/1412.6980.

[Kitao et al., 1991] Kitao, A., Hirata, F., and Gō, N. (1991). The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chemical physics*, 158(2-3):447–472. doi:10.1016/0301-0104(91)87082-7.

[Kmiecik et al., 2016] Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chemical reviews*, 116(14):7898–7936. doi:10.1021/acs.chemrev.6b00163.

[Koda and Saito, 2020] Koda, S.-i. and Saito, S. (2020). Sufficiency of unidirectional allostery in KaiC in generating the cyanobacterial circadian rhythm. *bioRxiv*. URL: https://www.biorxiv.org/content/biorxiv/early/2020/04/02/2020.04.01.021055.full.pdf, doi:10.1101/2020.04.01.021055.

[Kohnke et al., 2020] Kohnke, B., Ullmann, T. R., Beckmann, A., Kabadshow, I., Haensel, D., Morgenstern, L., Dobrev, P., Groenhof, G., Kutzner, C., Hess, B., et al. (2020). GROMEX: A Scalable and Versatile Fast Multipole Method for Biomolecular Simulation. In Bungartz, H.-J., Reiz, S., Uekermann, B., Neumann, P., and Nagel, W. E., editors, *Software for Exascale Computing-SPPEXA 2016-2019*, volume 136, pages 517–543. Springer. doi:10.1007/978-3-030-47956-5_17.

[Kondrashov et al., 2007] Kondrashov, D. A., Van Wynsberghe, A. W., Bannen, R. M., Cui, Q., and Phillips Jr, G. N. (2007). Protein structural variation in computational models and crystallographic data. *Structure*, 15(2):169–177. URL: http://www.cell.com/article/S0969212607000299/pdf, doi:10.1016/j.str.2007.05.001.

[Kong and Brooks III, 1996] Kong, X. and Brooks III, C. L. (1996). lambda-dynamics: A new approach to free energy calculations. *The Journal of chemical physics*, 105(6):2414–2423.

[Koshland Jr et al., 1966] Koshland Jr, D. E., Némethy, G., and Filmer, D. (1966). Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5(1):365–385. doi:10.1021/bi00865a047.

[Kremer and Lützen, 2013] Kremer, C. and Lützen, A. (2013). Artificial allosteric receptors. *Chemistry–A European Journal*, 19(20):6162–6196. doi:10.1002/chem.201203814.

[Kundu et al., 2002] Kundu, S., Melton, J. S., Sorensen, D. C., and Phillips Jr, G. N. (2002). Dynamics of proteins in crystals: comparison of experiment with simple models. *Biophysical journal*, 83(2):723–732. URL: http://www.cell.com/article/S000634950275203X/pdf, doi:10.1016/s0006-3495(02)75203-x.

[Laine et al., 2012] Laine, E., Auclair, C., and Tchertanov, L. (2012). Allosteric communication across the native and mutated KIT receptor tyrosine kinase. 8:e1002661. doi:10.1371/journal.pcbi.1002661.

[Lamm and Szabo, 1986] Lamm, G. and Szabo, A. (1986). Langevin modes of macromolecules. *The Journal of chemical physics*, 85(12):7334–7348. doi:10.1063/1.451373.

[Landau and Lifshitz, 1986] Landau, L. and Lifshitz, E. (1986). A. d. M. Kosevich and LP Pitaevskii, Theory of elasticity, 3rd English edn.

[Lapolla et al., 2021] Lapolla, A., Vossel, M., and Godec, A. (2021). Time-and ensemble-average statistical mechanics of the Gaussian network model. *Journal of Physics A: Mathematical and Theoretical*, 54(35):355601. doi:10.1088/1751-8121/ac16e4.

[Laskowski et al., 2009] Laskowski, R. A., Gerick, F., and Thornton, J. M. (2009). The structural basis of allosteric regulation in proteins. *FEBS Letters*, 583(11):1692–1698. URL: http://doi.wiley.com/10.1016/j.febslet.2009.03.019, doi:10.1016/j.febslet.2009.03.019.

[Lazaridis and Karplus, 1999] Lazaridis, T. and Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins: Structure, Function, and Bioinformatics*, 35(2):133–152. URL: https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/%28SICI%291097-0134%2819990501%2935%3A2%3C133%3A%3AAID-PROT1%3E3.0.CO%3B2-N, doi:10.1002/(sici)1097-0134(19990501)35:2<133::aid-prot1>3.0.co;2-n.

[Lerner et al., 2014] Lerner, E., DeGiuli, E., Düring, G., and Wyart, M. (2014). Breakdown of continuum elasticity in amorphous solids. *Soft Matter*, 10(28):5085–5092. URL: http://arxiv.org/pdf/1312.2146, doi:10.1039/c4sm00311j.

[Levitt, 1983] Levitt, M. (1983). Molecular dynamics of native protein. I. Computer simulation of trajectories. *Journal of molecular biology*, 168(3):595–617.

[Levitt et al., 1983] Levitt, M., Sander, C., and Stern, P. S. (1983). The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *International Journal of Quantum Chemistry*, 24(S10):181–199. doi:10.1002/qua.560240721.

[Levy et al., 1982] Levy, R. M., Perahia, D., and Karplus, M. (1982). Molecular dynamics of an alpha-helical polypeptide: temperature dependence and deviation from harmonic behavior. *Proceedings of the National Academy of Sciences*, 79(4):1346–1350. doi:10.1073/pnas.79.4.1346.

[Lezon et al., 2009] Lezon, T. R., Shrivastava, I. H., Yang, Z., and Bahar, I. (2009). *Elastic Network Models For Biomolecular Dynamics: Theory and Application to*

Membrane Proteins and Viruses, volume 10, page 129–158. WORLD SCIENTIFIC. URL: http://www.worldscientific.com/doi/abs/10.1142/9789812838803_0007, doi:10.1142/9789812838803_0007.

[Liang and Dill, 2001] Liang, J. and Dill, K. A. (2001). Are proteins well-packed? *Biophysical journal*, 81(2):751–766. URL: http://www.cell.com/article/S0006349501757396/pdf, doi:10.1016/s0006-3495(01)75739-6.

[Liang et al., 1998a] Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., and Subramaniam, S. (1998a). Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins: Structure, Function, and Bioinformatics*, 33(1):1–17. doi:10.1002/(sici)1097-0134(19981001)33:1<1::aid-prot1>3.0.co;2-o.

[Liang et al., 1998b] Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P. V., and Subramaniam, S. (1998b). Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins: Structure, Function, and Bioinformatics*, 33(1):18–29. doi:10.1002/(sici)1097-0134(19981001)33:1<18::aid-prot2>3.0.co;2-h.

[Light and Anderson, 2013] Light, S. H. and Anderson, W. F. (2013). The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis. *Protein Science*, 22(4):395–404. URL: https://europepmc.org/articles/pmc3610045?pdf=render, doi:10.1002/pro.2233.

[Liu, 2021] Liu, A. (2021). Doing "Statistical Mechanics" with Big Data: Understanding Protein Allostery. *Bulletin of the American Physical Society*. URL: https://meetings.aps.org/Meeting/MAR21/Session/S22.1.

[Liu et al., 2011] Liu, A. J., Nagel, S. R., Van Saarloos, W., and Wyart, M. (2011). The jamming scenario—an introduction and outlook. *Dynamical heterogeneities in glasses, colloids, and granular media*, page 298. URL: http://arxiv.org/pdf/1006.2365, doi:10.1093/acprof:oso/9780199691470.003.0009.

[Lopez-Blanco et al., 2013] Lopez-Blanco, J. R., Reyes, R., Aliaga, J. I., Badia, R. M., Chacon, P., and Quintana-Orti, E. S. (2013). Exploring large macromolecular functional motions on clusters of multicore processors. *Journal of Computational Physics*, 246:275–288. doi:10.1016/j.jcp.2013.03.032.

[López-Blanco et al., 2014] López-Blanco, J. R., Miyashita, O., Tama, F., and Chacón, P. (2014). Normal mode analysis techniques in structural biology. *Els*. doi:10.1002/9780470015902.a0020204.pub2.

[Ma et al., 2016] Ma, X., Meng, H., and Lai, L. (2016). Motions of allosteric and orthosteric ligand-binding sites in proteins are highly correlated. *Journal of Chemical Information and Modeling*, 56(9):1725–1733. doi:10.1021/acs.jcim.6b00039.

[Macarron et al., 2011] Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., et al. (2011). Impact of high-throughput screening in biomedical research. *Nature reviews Drug discovery*, 10(3):188–195. doi:10.1038/nrd3368.

[Marques and Sanejouand, 1995] Marques, O. and Sanejouand, Y.-H. (1995). Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins: Structure, Function, and Bioinformatics*, 23(4):557–560. doi:10.1002/prot.340230410.

[Maxwell, 1864] Maxwell, J. C. (1864). L. On the calculation of the equilibrium and stiffness of frames. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 27(182):294–299. URL: https://www.tandfonline.com/doi/full/10.1080/14786446408643668, doi:10.1080/14786446408643668.

[McCammon et al., 1977] McCammon, J. A., Gelin, B. R., and Karplus, M. (1977). Dynamics of folded proteins. *nature*, 267(5612):585–590. doi:10.1038/267585a0.

[McCammon et al., 1976] McCammon, J. A., Gelin, B. R., Karplus, M., and Wolynes, P. G. (1976). The hinge-bending mode in lysozyme. *Nature*, 262(5566):325–326. URL: https://www.nature.com/articles/262325a0, doi:10.1038/262325a0.

[McLeish et al., 2013] McLeish, T. C., Rodgers, T., and Wilson, M. R. (2013). Allostery without conformation change: modelling protein dynamics at multiple scales. *Physical biology*, 10(5):056004. URL: http://dro.dur.ac.uk/13295/1/13295.pdf, doi:10.1088/1478-3975/10/5/056004.

[Mitchell et al., 2016] Mitchell, M. R., Tlusty, T., and Leibler, S. (2016). Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings. *Proceedings of the National Academy of Sciences*, 113(40):E5847–E5855. URL: https://www.pnas.org/content/pnas/113/40/E5847.full.pdf, doi:10.1073/pnas.1609462113.

[Mitternacht and Berezovsky, 2011] Mitternacht, S. and Berezovsky, I. N. (2011). Coherent conformational degrees of freedom as a structural basis for allosteric communication. *PLoS computational biology*, 7(12):e1002301. doi:10.1371/journal.pcbi.1002301.

[Miyashita et al., 2003] Miyashita, O., Onuchic, J. N., and Wolynes, P. G. (2003). Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proceedings of the National Academy of Sciences*, 100(22):12570–12575. URL: http://www.pnas.org/content/100/22/12570.full.pdf, doi:10.1073/pnas.2135471100.

[Monod and Jacob, 1961] Monod, J. and Jacob, F. (1961). General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor symposia on quantitative biology*, volume 26, pages 389–401. Cold Spring Harbor Laboratory Press, Elsevier. doi:10.1016/b978-0-12-460482-7.50044-0.

[Monod et al., 1965] Monod, J., Wyman, J., and Changeux, J.-P. (1965). On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology*, 12(1):88–118. URL: https://www.sciencedirect.com/science/article/pii/S0022283665802856, doi:https://doi.org/10.1016/S0022-2836(65)80285-6.

[Motlagh et al., 2014] Motlagh, H. N., Wrabl, J. O., Li, J., and Hilser, V. J. (2014). The ensemble nature of allostery. *Nature*, 508(7496):331–339. URL: https://europepmc.org/articles/pmc4224315?pdf=render, doi:10.1038/nature13001.

[Müller et al., 1996] Müller, C., Schlauderer, G., Reinstein, J., and Schulz, G. E. (1996). Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure*, 4(2):147–156. URL: http://www.cell.com/article/S0969212696000184/pdf, doi:10.1016/s0969-2126(96)00018-4.

[Müller and Schulz, 1992] Müller, C. W. and Schulz, G. E. (1992). Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 Å resolution: A model for a catalytic transition state. *Journal of molecular biology*, 224(1):159–177.

[Na et al., 2018] Na, H., Ben-Avraham, D., and Tirion, M. M. (2018). Slow normal modes of proteins are accurately reproduced across different platforms. *Physical Biology*, 16(1):016003. URL: http://arxiv.org/pdf/1806.10656, doi:10.1088/1478-3975/aae333.

[Neuroscience, 2008] Neuroscience, F. E. (2008). Dale Purves, George J. *Augustine, David Fitzpatrick, William C. Hall, Anthony-Samuel LaMantia, James O. McNamara, and Leonard E. White, Sinauer*.

[Ni et al., 2021a] Ni, D., Chai, Z., Wang, Y., Li, M., Yu, Z., Liu, Y., Lu, S., and Zhang, J. (2021a). Along the allostery stream: Recent advances in computational methods

for allosteric drug discovery. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, page e1585. doi:10.1002/wcms.1585.

[Ni et al., 2021b] Ni, D., Wei, J., He, X., Rehman, A. U., Li, X., Qiu, Y., Pu, J., Lu, S., and Zhang, J. (2021b). Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. *Chemical science*, 12(1):464–476. URL: https://pubs.rsc.org/en/content/articlepdf/2021/sc/d0sc05131d, doi:10.1039/d0sc05131d.

[Nicolay and Sanejouand, 2006] Nicolay, S. and Sanejouand, Y.-H. (2006). Functional Modes of Proteins Are among the Most Robust. *Phys. Rev. Lett.*, 96(7):078104. Publisher: American Physical Society. URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.078104, doi:10.1103/PhysRevLett.96.078104.

[Noguti and Gō, 1982] Noguti, T. and Gō, N. (1982). Collective variable description of small-amplitude conformational fluctuations in a globular protein. *Nature*, 296(5859):776–778. doi:10.1038/296776a0.

[Nussinov and Tsai, 2013] Nussinov, R. and Tsai, C.-J. (2013). Allostery in Disease and in Drug Discovery. *Cell*, 153(2):293–305. URL: https://linkinghub.elsevier.com/retrieve/pii/S0092867413003917, doi:10.1016/j.cell.2013.03.034.

[Okazaki and Takada, 2008] Okazaki, K.-i. and Takada, S. (2008). Dynamic energy landscape view of coupled binding and protein conformational change: induced-fit versus population-shift mechanisms. *Proceedings of the National Academy of Sciences*, 105(32):11182–11187. URL: https://europepmc.org/articles/pmc2516237?pdf=render, doi:10.1073/pnas.0802524105.

[Olaussen and Stell, 1991] Olaussen, K. and Stell, G. (1991). New microscopic approach to the statistical mechanics of chemical association. *Journal of statistical physics*, 62(1):221–237. doi:10.1007/bf01020867.

[Papoulis and Pillai, 2002] Papoulis, A. and Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*, volume 79. Tata McGraw-Hill Education. doi:10.2307/2288754.

[Park et al., 2004] Park, S. H., Goo, J. M., and Jo, C.-H. (2004). Receiver operating characteristic (ROC) curve: practical review for radiologists. *Korean journal of radiology*, 5(1):11–18. URL: https://europepmc.org/articles/pmc2698108?pdf=render, doi:10.3348/kjr.2004.5.1.11.

[Pellegrino and Calladine, 1986] Pellegrino, S. and Calladine, C. (1986). Matrix analysis of statically and kinematically indeterminate frameworks. *International Journal of Solids and Structures*, 22(4):409–428. URL: https://linkinghub.elsevier.com/retrieve/pii/0020768386900144, doi:10.1016/0020-7683(86)90014-4.

[Perahia and Mouawad, 1995] Perahia, D. and Mouawad, L. (1995). Computation of low-frequency normal modes in macromolecules: improvements to the method of diagonalization in a mixed basis and application to hemoglobin. *Computers and chemistry*, 19(3):241–246. doi:10.1016/0097-8485(95)00011-g.

[Perutz, 1970] Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin: haem–haem interaction and the problem of allostery. *Nature*, 228(5273):726–734. doi:10.1038/228726a0.

[Perutz et al., 1960] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. (1960). Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature*, 185(4711):416–422. doi:10.1038/185416a0.

[Peselis et al., 2015] Peselis, A., Gao, A., and Serganov, A. (2015). Cooperativity, allostery and synergism in ligand binding to riboswitches. *Biochimie*, 117:100–109. URL: https://europepmc.org/articles/pmc4643686?pdf=render, doi:10.1016/j.biochi.2015.06.028.

[Peselis and Serganov, 2021] Peselis, A. and Serganov, A. (2021). Cooperativity and Allostery in RNA Systems. *Allostery*, pages 255–271.

[Poma et al., 2018] Poma, A. B., Li, M. S., and Theodorakis, P. E. (2018). Generalization of the elastic network model for the study of large conformational changes in biomolecules. *Physical Chemistry Chemical Physics*, 20(25):17020–17028. URL: http://xlink.rsc.org/?DOI=C8CP03086C, doi:10.1039/C8CP03086C.

[Pronk et al., 2013] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., et al. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854. URL: https://academic.oup.com/bioinformatics/article-pdf/29/7/845/17343875/btt055.pdf, doi:10.1093/bioinformatics/btt055.

[Purcell, 1977] Purcell, E. M. (1977). Life at low Reynolds number. *American journal of physics*, 45(1):3–11. doi:10.1119/1.10903.

[Rafferty et al., 1989] Rafferty, J. B., Somers, W. S., Saint-Girons, I., and Phillips, S. E. (1989). Three-dimensional crystal structures of Escherichia coli met repressor with and without corepressor. *Nature*, 341(6244):705–710. doi:10.1038/341705a0.

[Ravasio, 2020] Ravasio, R. (2020). *Mechanics and co-evolution of allosteric materials and proteins*. PhD thesis, EPFL.

[Ravasio et al., 2019] Ravasio, R., Flatt, S., Yan, L., Zamuner, S., Brito, C., and Wyart, M. (2019). Mechanics of allostery: contrasting the induced fit and population shift scenarios. *Biophysical Journal*, 117:S0006349519308306. arXiv: 1906.05043. URL: http://www.cell.com/article/S0006349519308306/pdf, doi:10.1016/j.bpj.2019.10.002.

[Rocks, 2019] Rocks, J. W. (2019). *Allosteric Functionality In Mechanical And Flow Networks*. PhD thesis, University of Pennsylvania.

[Rocks et al., 2020] Rocks, J. W., Liu, A. J., and Katifori, E. (2020). Revealing structure-function relationships in functional flow networks via persistent homology. *Physical Review Research*, 2(3):033234. URL: http://link.aps.org/pdf/10.1103/PhysRevResearch.2.033234, doi:10.1103/PhysRevResearch.2.033234.

[Rocks et al., 2021] Rocks, J. W., Liu, A. J., and Katifori, E. (2021). Hidden Topological Structure of Flow Network Functionality. *Physical Review Letters*, 126(2):028102. URL: http://arxiv.org/pdf/1911.11606, doi:10.1103/PhysRevLett.126.028102.

[Rocks et al., 2017] 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*, 114(10):2520–2525. URL: https://www.pnas.org/content/pnas/114/10/2520.full.pdf, doi:10.1073/pnas.1612139114.

[Rocks et al., 2019] Rocks, J. W., Ronellenfitsch, H., Liu, A. J., Nagel, S. R., and Katifori, E. (2019). Limits of multifunctionality in tunable networks. *Proceedings of the National Academy of Sciences*, 116(7):2506–2511. URL: https://www.pnas.org/content/pnas/116/7/2506.full.pdf, doi:10.1073/pnas.1806790116.

[Rodriguez et al., 2010] Rodriguez, G. J., Yao, R., Lichtarge, O., and Wensel, T. G. (2010). Evolution-guided discovery and recoding of allosteric pathway specificity determinants in psychoactive bioamine receptors. *Proceedings of the National Academy of Sciences*, 107(17):7787–7792. URL: http://www.pnas.org/content/107/17/7787.full.pdf, doi:10.1073/pnas.0914877107.

[Rouse, 1953] Rouse, P. E. (1953). A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers. *J. Chem. Phys.*, 21(7):1272–1280. URL: http://aip.scitation.org/doi/10.1063/1.1699180, doi:10.1063/1.1699180.

[Sacquin-Mora et al., 2007] Sacquin-Mora, S., Laforet, E., and Lavery, R. (2007). Locating the active sites of enzymes using mechanical properties. *Proteins: Structure, Function, and Bioinformatics*, 67(2):350–359. doi:10.1002/prot.21353.

[Sanejouand, 2013] Sanejouand, Y.-H. (2013). Elastic Network Models: Theoretical and Empirical Foundations. In Monticelli, L. and Salonen, E., editors, *Biomolecular Simulations: Methods and Protocols*, page 601–616. Humana Press. doi:10.1007/978-1-62703-017-5_23.

[Schlessinger, 1986] Schlessinger, J. (1986). Allosteric regulation of the epidermal growth factor receptor kinase. *The Journal of cell biology*, 103(6 Pt 1):2067–2072. URL: https://rupress.org/jcb/article-pdf/103/6/2067/1054121/2067.pdf, doi:10.1083/jcb.103.6.2067.

[Schumacher et al., 1997] Schumacher, M. A., Glasfeld, A., Zalkin, H., and Brennan, R. G. (1997). The X-ray structure of the PurR-guanine-purF operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity. *Journal of Biological Chemistry*, 272(36):22648–22653. URL: http://www.jbc.org/content/272/36/22648.full.pdf, doi:10.1074/jbc.272.36.22648.

[Selvaratnam et al., 2011] Selvaratnam, R., Chowdhury, S., VanSchouwen, B., and Melacini, G. (2011). Mapping allostery through the covariance analysis of NMR chemical shifts. *Proceedings of the National Academy of Sciences*, 108(15):6133–6138. URL: http://www.pnas.org/content/108/15/6133.full.pdf, doi:10.1073/pnas.1017311108.

[Shaw et al., 2010] Shaw, D. E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R. O., Eastwood, M. P., Bank, J. A., Jumper, J. M., Salmon, J. K., Shan, Y., et al. (2010). Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346. doi:10.1126/science.1187409.

[Shiau et al., 2006] Shiau, A. K., Harris, S. F., Southworth, D. R., and Agard, D. A. (2006). Structural analysis of E. coli hsp90 reveals dramatic nucleotide-dependent conformational rearrangements. *Cell*, 127(2):329–340. URL: http://www.cell.com/article/S009286740601275X/pdf, doi:10.1016/j.cell.2006.09.027.

[Shrake and Rupley, 1973] Shrake, A. and Rupley, J. A. (1973). Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of molecular biology*, 79(2):351–371. doi:10.1016/0022-2836(73)90011-9.

[Sierra et al., 2005] Sierra, D. P., Weir, N. A., and Jones, J. F. (2005). *A review of research in the field of nanorobotics.* Sandia National Laboratories. URL: http://www.osti.gov/servlets/purl/875622-4rE04b/, doi:10.2172/875622.

[Sievers and Higgins, 2014] Sievers, F. and Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. In *Multiple sequence alignment methods*, pages 105–116. Springer. doi:10.1007/978-1-62703-646-7_6.

[Smirnov, 1948] Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *The annals of mathematical statistics*, 19(2):279–281. doi:10.1214/aoms/1177730256.

[Smith et al., 2016] Smith, C. A., Ban, D., Pratihar, S., Giller, K., Paulat, M., Becker, S., Griesinger, C., Lee, D., and de Groot, B. L. (2016). Allosteric switch regulates protein–protein binding through collective motion. *Proceedings of the National Academy of Sciences*, 113(12):3269–3274. URL: https://www.pnas.org/content/pnas/113/12/3269.full.pdf, doi:10.1073/pnas.1519609113.

[Strang, 1993] Strang, G. (1993). The fundamental theorem of linear algebra. *The American Mathematical Monthly*, 100(9):848–855. doi:10.2307/2324660.

[Su et al., 2008] Su, J. G., Li, C. H., Hao, R., Chen, W. Z., and Xin Wang, C. (2008). Protein Unfolding Behavior Studied by Elastic Network Model. *Biophysical Journal*, 94(12):4586–4596. URL: http://www.cell.com/article/S0006349508703266/pdf, doi:10.1529/biophysj.107.121665.

[Suhre and Sanejouand, 2004] Suhre, K. and Sanejouand, Y.-H. (2004). ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic acids research*, 32(supp 2):W610–W614. doi:10.1093/nar/gkh368.

[Tama and Brooks III, 2005] Tama, F. and Brooks III, C. L. (2005). Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis. *Journal of molecular biology*, 345(2):299–314. doi:10.1016/j.jmb.2004.10.054.

[Tama and Sanejouand, 2001] Tama, F. and Sanejouand, Y.-H. (2001). Conformational change of proteins arising from normal mode calculations. *Protein engineering*, 14(1):1–

6. URL: https://academic.oup.com/peds/article-pdf/14/1/1/18544720/140001.pdf, doi:10.1093/protein/14.1.1.

[Tarnai, 1980] Tarnai, T. (1980). Simultaneous static and kinematic indeterminacy of space trusses with cyclic symmetry. *International Journal of Solids and Structures*, 16(4):347–359. doi:10.1016/0020-7683(80)90087-6.

[Taylor et al., 2014] Taylor, D., Cawley, G., and Hayward, S. (2014). Quantitative method for the assignment of hinge and shear mechanism in protein domain movements. *Bioinformatics*, 30(22):3189–3196. URL: https://academic.oup.com/bioinformatics/article-pdf/30/22/3189/17143978/btu506.pdf, doi:10.1093/bioinformatics/btu506.

[Tee et al., 2018] 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(6):e1006228. URL: https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1006228&type=printable, doi:10.1371/journal.pcbi.1006228.

[Tekpinar, 2018] Tekpinar, M. (2018). Flexible fitting to cryo-electron microscopy maps with coarse-grained elastic network models. *Molecular Simulation*, 44(8):688–696. doi:10.1080/08927022.2018.1431835.

[Teşileanu et al., 2015] Teşileanu, T., Colwell, L. J., and Leibler, S. (2015). Protein sectors: statistical coupling analysis versus conservation. *PLoS computational biology*, 11(2):e1004091. doi:10.1371/journal.pcbi.1004091.

[Thirumalai et al., 2019] Thirumalai, D., Hyeon, C., Zhuravlev, P. I., and Lorimer, G. H. (2019). Symmetry, rigidity, and allosteric signaling: from monomeric proteins to molecular machines. *Chemical reviews*, 119(12):6788–6821. URL: http://arxiv.org/pdf/1812.04969, doi:10.1021/acs.chemrev.8b00760.

[Tian et al., 2018] Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, 46(W1):W363–W367. doi:10.1093/nar/gky473.

[Tian et al., 2010] Tian, Y., Shirinzadeh, B., and Zhang, D. (2010). Design and dynamics of a 3-DOF flexure-based parallel mechanism for micro/nano manipulation. *Microelectron Eng*, 87(2):230–241. URL: https://www.sciencedirect.com/science/article/abs/pii/S0167931709005255?via%3Dihub, doi:10.1016/j.mee.2009.08.001.

[Tirion, 1996] Tirion, M. M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical Review Letters*, 77(9):1905–1908. doi:10.1103/PhysRevLett.77.1905.

[Togashi and Flechsig, 2018] Togashi, Y. and Flechsig, H. (2018). Coarse-Grained Protein Dynamics Studies Using Elastic Network Models. *Int. J. Mol. Sci.*, 19:18. URL: https://www.mdpi.com/1422-0067/19/12/3899/pdf, doi:10.3390/ijms19123899.

[Togashi and Mikhailov, 2007] Togashi, Y. and Mikhailov, A. S. (2007). Nonlinear relaxation dynamics in elastic networks and design principles of molecular machines. *Proceedings of the National Academy of Sciences*, 104(21):8697–8702. URL: https://europepmc.org/articles/pmc1868896?pdf=render, doi:10.1073/pnas.0702950104.

[Togashi et al., 2010] Togashi, Y., Yanagida, T., and Mikhailov, A. S. (2010). Nonlinearity of mechanochemical motions in motor proteins. *PLoS computational biology*, 6(6):e1000814. doi:10.1371/journal.pcbi.1000814.

[Tsai and Nussinov, 2014] Tsai, C.-J. and Nussinov, R. (2014). A Unified View of "How Allostery Works". *PLoS Computational Biology*, 10(2):e1003394. URL: http://dx.plos.org/10.1371/journal.pcbi.1003394, doi:10.1371/journal.pcbi.1003394.

[Umbarger, 1956] Umbarger, H. E. (1956). Evidence for a negative-feedback mechanism in the biosynthesis of isoleucine. *Science*, 123(3202):848–848. doi:10.1126/science.123.3202.848-a.

[Van Der Spoel et al., 2005] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26(16):1701–1718. doi:10.1002/jcc.20291.

[VanWart et al., 2012] VanWart, A. T., Eargle, J., Luthey-Schulten, Z., and Amaro, R. E. (2012). Exploring residue component contributions to dynamical network models of allostery. *Journal of chemical theory and computation*, 8(8):2949–2961. URL: https://europepmc.org/articles/pmc3489502?pdf=render, doi:10.1021/ct300377a.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272. URL: http://arxiv.org/abs/1907.10121, doi:10.1038/s41592-019-0686-2.

[Walker et al., 2020] Walker, A. S., Russ, W. P., Ranganathan, R., and Schepartz, A. (2020). RNA sectors and allosteric function within the ribosome. *Proceedings of the National Academy of Sciences*, 117(33):19879–19887. URL: https://www.pnas.org/content/pnas/117/33/19879.full.pdf, doi:10.1073/pnas.1909634117.

[Wand, 2013] Wand, A. J. (2013). The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Current opinion in structural biology*, 23(1):75–81. URL: https://europepmc.org/articles/pmc3572299?pdf=render, doi:10.1016/j.sbi.2012.11.005.

[Wardell et al., 2002] Wardell, M., Wang, Z., Ho, J. X., Robert, J., Ruker, F., Ruble, J., and Carter, D. C. (2002). The atomic structure of human methemalbumin at 1.9 Å. *Biochemical and biophysical research communications*, 291(4):813–819. doi:10.1006/bbrc.2002.6540.

[Weiner et al., 1984] Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984). A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, 106(3):765–784. doi:10.1021/ja00315a051.

[Weinkam et al., 2013] Weinkam, P., Chen, Y. C., Pons, J., and Sali, A. (2013). Impact of mutations on the allosteric conformational equilibrium. *Journal of molecular biology*, 425(3):647–661. URL: https://europepmc.org/articles/pmc3557769?pdf=render, doi:10.1016/j.jmb.2012.11.041.

[Wenthur et al., 2014] Wenthur, C. J., Gentry, P. R., Mathews, T. P., and Lindsley, C. W. (2014). Drugs for allosteric sites on receptors. *Annual review of pharmacology and toxicology*, 54:165–184. URL: https://europepmc.org/articles/pmc4063350?pdf=render, doi:10.1146/annurev-pharmtox-010611-134525.

[Williams and Hall, 2014] Williams, S. G. and Hall, K. B. (2014). Linkage and allostery in snRNP protein/RNA complexes. *Biochemistry*, 53(22):3529–3539. doi:10.1021/bi500192a.

[Wu and Lander, 2020] Wu, M. and Lander, G. C. (2020). How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Current opinion in structural biology*, 64:9–16. URL: https://europepmc.org/articles/pmc7666008?pdf=render, doi:10.1016/j.sbi.2020.05.007.

[Wyart, 2005] Wyart, M. (2005). On the rigidity of amorphous solids. *Ann. Phys. Fr.*, 30(3):1–96. doi:10.1051/anphys:2006003.

[Xu et al., 2003] Xu, C., Tobi, D., and Bahar, I. (2003). Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T–R2 transition. *Journal of molecular biology*, 333(1):153–168. doi:10.1016/j.jmb.2003.08.027.

[Yan et al., 2017] Yan, L., Ravasio, R., Brito, C., and Wyart, M. (2017). Architecture and coevolution of allosteric materials. *Proceedings of the National Academy of Sciences*, 114(10):2526–2531. URL: https://www.pnas.org/content/pnas/114/10/2526.full.pdf, doi:10.1073/pnas.1615536114.

[Yan et al., 2018] Yan, L., Ravasio, R., Brito, C., and Wyart, M. (2018). Principles for Optimal Cooperativity in Allosteric Materials. *Biophysical Journal*, 114(12):2787–2798. URL: http://www.cell.com/article/S0006349518305939/pdf, doi:10.1016/j.bpj.2018.05.015.

[Yang et al., 1997] Yang, C., Drakos, N., Lehoucq, R. B., and Sorensen, D. C. (1997). ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods. URL: https://www.caam.rice.edu/software/ARPACK/UG/node33.html.

[Yang and Bahar, 2005] Yang, L.-W. and Bahar, I. (2005). Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. *Structure*, 13(6):893–904. URL: http://www.cell.com/article/S096921260500167X/pdf, doi:10.1016/j.str.2005.03.015.

[Yang et al., 2009] Yang, L.-W., Eyal, E., Bahar, I., and Kitao, A. (2009). Principal component analysis of native ensembles of biomolecular structures (PCA NEST): insights into functional dynamics. *Bioinformatics*, 25(5):606–614. URL: https://academic.oup.com/bioinformatics/article-pdf/25/5/606/16891522/btp023.pdf, doi:10.1093/bioinformatics/btp396.

[Yates and Pardee, 1956] Yates, R. A. and Pardee, A. B. (1956). Control of pyrimidine biosynthesis in Escherichia coli by a feed-back mechanism. *Journal of Biological Chemistry*, 221(2):757–770. doi:10.1016/s0021-9258(18)65188-9.

[Ye et al., 2006] Ye, J., McGinnis, S., and Madden, T. L. (2006). BLAST: improvements for better sequence analysis. *Nucleic acids research*, 34(suppl 2):W6–W9. doi:10.1093/nar/gkl164.

[Yuan et al., 2003] Yuan, Z., Zhao, J., and Wang, Z.-X. (2003). Flexibility analysis of enzyme active sites by crystallographic temperature factors. *Protein Engineering*, 16(2):109–114. URL: https://academic.oup.com/peds/article-pdf/16/2/109/18547149/160109.pdf, doi:10.1093/proeng/gzg014.

[Zhang and Bishop, 2007] Zhang, X.-Y. and Bishop, A. C. (2007). Site-specific incorporation of allosteric-inhibition sites in a protein tyrosine phosphatase. *Journal of the American Chemical Society*, 129(13):3812–3813. URL: https://europepmc.org/articles/pmc2518263?pdf=render, doi:10.1021/ja069098t.

[Zheng, 2011] Zheng, W. (2011). Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophysical journal*, 100(2):478–488. URL: http://www.cell.com/article/S0006349510034636/pdf, doi:10.1016/j.bpj.2010.12.3680.

[Zheng and Brooks, 2005a] Zheng, W. and Brooks, B. (2005a). Identification of Dynamical Correlations within the Myosin Motor Domain by the Normal Mode Analysis of an Elastic Network Model. *J. Mol. Bio.*, 346(3):745–759. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022283604015906, doi:10.1016/j.jmb.2004.12.020.

[Zheng and Brooks, 2005b] Zheng, W. and Brooks, B. R. (2005b). Probing the Local Dynamics of Nucleotide-Binding Pocket Coupled to the Global Dynamics: Myosin versus Kinesin. *Biophysical Journal*, 89(1):167–178. URL: https://linkinghub.elsevier.com/retrieve/pii/S0006349505726680, doi:10.1529/biophysj.105.063305.

[Zheng et al., 2006] Zheng, W., Brooks, B. R., and Thirumalai, D. (2006). Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proceedings of the National Academy of Sciences*, 103(20):7664–7669. URL: http://www.pnas.org/content/103/20/7664.full.pdf, doi:10.1073/pnas.0510426103.

[Zheng and Doniach, 2003] Zheng, W. and Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. *Proceedings of the National Academy of Sciences*, 100(23):13253–13258. URL: http://www.pnas.org/content/100/23/13253.full.pdf, doi:10.1073/pnas.2235686100.