Coarse-Grained Models of Biomolecule Dynamics and Allostery

by

Hedvika Toncrova

Submitted in accordance with the requirements for the
degree of Doctor of Philosophy

in the
Faculty of Mathematics and Physical Sciences
School of Physics and Astronomy

December 2010

Declaration of Authorship

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Abstract

Recently, it has become increasingly accepted that thermal fluctuations take active part in functional tasks of biological molecules. We employ a set of coarse-grained models to investigate the mechanism of transmission of allosteric signal via spatial fluctuations. Our models are coarser than those in computational techniques established in molecular biology, but allow for both the identification of candidates for the essential physical structures and also the analytical determination of thermodynamic quantities that define ligand binding. The models are constructed for general classes of macromolecules and are validated through parameterisation from experiments and atomistic simulations.

In the first part of this thesis we investigate the "dynamic allostery" in dimeric proteins composed of two identical subunits. We demonstrate that cooperative effects upon binding of two identical ligands can arise purely through modification of slow global vibrational modes of the protein. We parameterise the model on a test case, the CAP homodimer. Finally, we explain the role of local, fast vibrations in the allosteric effect and propose a general protocol for interpreting thermodynamic parameters of dynamically allosteric homodimers.

The second part of this thesis considers allosteric effects in DNA, an example of nearly uniform elastic medium. The DNA is modeled as an elastic rod and substrate binding as local increase of its bending and twisting rigidity. This results in altered structure of normal modes and leads to qualitatively different type of dynamic allostery compared to that of the discrete models previously employed to study allosteric effects in proteins. Dynamic allostery in DNA is found always to be negative, due to an anti-correlated amplitude of thermal fluctuations at the binding site and around it. This allows us to draw conclusions about general design rules of allosteric molecules and highlight the controlling feature that biological molecules evolved to optimize their dynamics for their function.

Acknowledgements

Contents

List of Figures

List of Tables

Abbreviations

bp	base pair
cAMP	Cyclic Adenosine Monophosphate
CAP	Catabolite Activator Protein
CTP	cytidine-5'-triphosphate
DNA	Deoxyribonucleic Acid
DPG	2,3-Diphophoglycerate
ENM	Elastic Network Model
FRET	Fluorescence Resonance Energy Transfer
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GCT	glycerol-3-phosphate cytidylyltransferase
GNM	Gaussian Network Model
H33258	Hoechst 33258
Hb	Haemoglobin
ITC	Isothermal Titration Calorimetry
KNF	Koshland Nemethy Filmer
MD	Molecular Dynamics
MWC	Monod Wyman Changeux
NMA	Normal Mode Analysis
NMR	Nuclear Magnetic Resonance
PCA	Principal Component Analysis
RTB	Rotation Translation Block

To my family / Mé Rodině.

Chapter 1

Introduction

Proteins, frequently dubbed the building blocks of life, are large molecules involved in nearly every cellular process. The vast majority of proteins in their functional state are folded into complicated, yet delicate and precisely defined three dimensional structures. This has been recognised since John C. Kendrew and Max F. Perutz resolved the structure of myoglobin in 1958 [1]. Not long after their discovery it became clear that proteins constantly unfold and refold, but critically they can function fully only in the folded state. This led to the hypothesis that function is determined by the structure, widely accepted until recently.

The hypothesis has been substantially supported by the privileged position and great success of X-ray crystallography in molecular biology. The technique, pioneered by Kendrew and Perutz, measures the average positions of each atom in the protein and produces static three dimensional structures. For a long time scientists focused on creating a database of all protein structures believing that together with a description of reaction pathways it would constitute a complete understanding of molecular biology.

Confidence in this approach began to be shaken during the 1970s when biophysicists first pointed out the importance of dynamics in function [2]. Three decades later, it has become an accepted concept that dynamics, or thermally excited vibrations, of the protein are just as important to its function as is the structure. Furthermore, the "new view" suggests that proteins have evolved to harness thermal motion along with its structure to fulfill their role within the cell [3]. Recent publications have highlighted the key role of protein dynamics in many aspects of protein function: intramolecular signal

transmission [4, 5], inter-protein signaling [6, 7], ligand recognition [8], enzyme catalysis [9], protein folding [10], complex assembly [7], chaperonin assisted protein folding [11], ion channel selectivity [12] and ATP synthesis [13]. In this thesis we study the role of thermal motions in intramolecular signal transmission, commonly referred to as allostery.

Biological systems are complex networks whose operation is achieved through an accurate regulation. Proteins, the main constituents of the network, play their part by participating in concerted sequences of binding events. These sequences often involve binding of multiple ligands to the protein's distinct binding sites. By a ligand we mean any molecule that associates with the protein in question. The size of a ligand can vary dramatically - ranging from a single ion, through a small organic compound to another protein or DNA. In most cases the binding events are not independent, binding of one ligand influences the affinity for the second ligand - a phenomenon known as allostery.

Allostery is a very intriguing effect, not only because it allows proteins to assume the role of components in logical circuits, but also because it generally involves a long-range communication through the protein body. The binding sites can be more than a hundred Ångströms apart, separated by a jungle of protein folds and/or solvent molecules [11]. The mechanism of the communication has long been investigated but many aspects still remain unclear. In accordance with the above-introduced theory, experts used to associate allostery purely with changes in the protein structure. Under that hypothesis the ligand binding must initiate some type of conformational change in the protein, such as a pocket opening, which in turn modifies the affinity for the second ligand. This may be the primary mechanism for many allosteric systems but as we know now it is not for all. Allosteric binding, like all chemical reactions is a complex thermodynamic event in which structural and dynamical changes contribute equally. In a more formal language, the equilibrium of an allosteric reaction is determined by the free energy of binding which is composed of an enthalpic and entropic term

$$\Delta G = \Delta H - T\Delta S. \tag{1.1}$$

The enthalpic term corresponds, in the first approximation, to changes in structure and the entropic to changes in dynamics.

The allosteric reactions are assumed to occur under constant pressure and the volume changes associated with them are assumed to be negligible. This allows us to use the Gibbs and Helmholtz free energy interchangeably. Because the Gibbs free energy is

standard in biochemistry and in the field of allostery, we denote the free energy by G in the whole thesis.

The significance of dynamics in allosteric events is generally acknowledged but the detailed mechanism that proteins use to take advantage of thermal motion is disputed [7]. Several lines of thought, hypothesising the mechanism, have emerged recently. One of them suggests that dynamics provides a route to conformational transition [14, 15]. The protein uses its thermal energy to explore its conformational space or free energy landscape. Ligand binding can alter this landscape resulting in different preferred conformations with modified affinity for the second ligand. An alternative line of thought is that only a small fraction of the protein atoms or residues, directly linking the binding sites, is involved in the propagation [16, 17].

In this thesis we investigate an alternative mechanism whereby a change in average position of atoms in the protein is not required and the signal is propagated through change in the average fluctuation amplitude. The main difference from the linked residue model is that a large proportion of the atoms is involved in the propagation. The information is communicated mainly through changes in low frequency, global, modes that move large protein domains in a concerted fashion. This alternative mechanism was suggested nearly 30 years ago by Cooper and Dryden who also coined the term dynamic allostery for it [18]. They, however, have not expanded on their original paper and their ideas remained purely theoretical. Beginning in 2004, Hawkins and McLeish produced a series of studies in which they developed Cooper and Dryden's ideas into concrete models of particular classes of proteins.

It is known that low frequency, global motions are the most relevant motions for allostery [19]. These modes lie currently, and in the foreseeable future, out of reach of routine atomistic simulations. In order to evaluate the contribution of the slow modes to dynamic allostery, Hawkins and McLeish had to identify the appropriate level of coarse-graining. They chose a very crude level, which nevertheless captures the desired effect. In their models whole domains or subdomains are treated as rigid bodies whose motion is governed by a set of harmonic potentials. They applied this concept to two diverse classes of proteins: the rigid dimers, a structure often adopted by DNA binding proteins, and coiled-coils seen commonly amongst molecular motors.

In this thesis we draw on their efforts with the aim to increase the arsenal of models applicable to real systems. This is a part of the long term goal to develop a systematic

coarse-graining method able to explain and/or predict allostery in an arbitrary system. We first focus on a model of proteins with two symmetrical binding sites (homodimers). After a thorough study of this system and illustration of the model's applicability on an example protein (CAP) we suggest an expansion to a general oligomer (protein with multiple binding sites).

The second part of the thesis addresses dynamic allostery in a somewhat unusual allosteric system - DNA. Allostery as a concept is usually presented in biochemistry textbooks as a fundamental regulation mechanism of proteins, although it is well known that DNA too can act as a mediator of allosteric signals [20]. This role of DNA has always been considered secondary and has received less attention. DNA's role as a medium for allosteric signals is in my opinion underestimated, as it plays part in important processes, such as regulation of gene expression or DNA crowding through unspecific binding related to the much debated chromatin properties [21]. DNA is also a common target of pharmaceutical research and our findings may have implications for the vibrant field of drug discovery.

DNA constitutes an ideal model system because of its well known mechanical properties and relatively simple structure. This enables us to use coarse-grained models that have been many times tested experimentally and to parameterise them from molecular dynamics simulations.

In the remainder of this chapter we introduce the basic thermodynamic principles of ligand binding. We explain where the protein dynamics come into picture and why dynamics cannot be discounted in the considerations of the cooperative binding. We expand on the topic of protein dynamics in general and list several experimental techniques capable to directly measure this effect.

In the following chapter we address the philosophy and the hierarchy of coarse-graining. We summarize several basic approaches and highlight those employed in this thesis. We finish by introducing and justifying our coarse-graining methodology. Three results chapters that follow focus on the cases of symmetric homodimer and DNA in detail. Each of these chapters begins with a detailed introduction to the particular system, the reasons for choosing it and the justification of the method of coarse-graining. Finally the last chapter summarizes the findings of this work and concludes with future plans.

In broader context, the aim of this work is to demonstrate that both structure and dynamics of a protein carry information on its function, and that mathematical modeling

in combination with physical laws can be useful when addressing questions in biology. We recognize that it is important to convey our message in such a fashion that the benefits of the theoretical approach and of the resulting models become obvious to wider biological and scientific circles. We therefore try to explain them in as simple and understandable terms as possible.

1.1 Thermodynamics of Ligand Binding

The function of the vast majority of biological molecules rests on their interaction with their environment. This calls, in turn, for selective binding of molecules of different types and sizes. If the selectivity or sensitivity for a particular molecule is disrupted the consequences can be severe, ranging from minor dysfunctions to the cell death. Understanding the way molecules interact is therefore of crucial importance and has been widely explored for more than a century.

For the sake of simplicity let us focus the following discussion on the thermodynamics of binding to a protein. The same conclusions apply to all macromolecules, in particular DNA.

Molecules that commonly associate with a protein range from single ions or small organic molecules to large molecules such as proteins and DNA. Irrespective of its size and type we refer to the molecule interacting with the protein as a ligand. The protein with the ligand bound is known as the holo form, the unbound state as the apo form.

The reaction between the protein and its ligand is determined by the affinity of the two for each other. High affinity signifies that the protein is likely to be in its holo form and vice versa. The measure of affinity is provided by the association constant K_a. The equilibrium state of a reaction of a protein P with a ligand L is described by the equation

$$P + L \rightleftharpoons P \cdot L \tag{1.2}$$

with the double ended arrow signifying equilibrium. The affinity is usually expressed in the form of the association constant K_a

$$K_a = \frac{[P \cdot L]}{[P][L]}, \tag{1.3}$$

where [P] is concentration of P. The greater association constant corresponds to a greater affinity. The association constant has units of $(\text{concentration})^{-1}$. The term constant may be slightly misleading, since all association constants depend on temperature and pressure (or volume). In literature the reciprocal of association constant, the dissociation constant, is sometimes used. The dissociation constant K_d has units of concentration and is thus for many more intuitive.

The association constant can also be expressed in terms of the rate constants k_a and k_d defined as

$$P + L \underset{k_d}{\overset{k_a}{\rightleftharpoons}} P \cdot L. \tag{1.4}$$

Then

$$K_a = \frac{k_a}{k_d}. \tag{1.5}$$

Another useful measure of affinity is the dimensionless saturation fraction Y

$$Y = \frac{[P \cdot L]}{[P] + [P \cdot L]}. \tag{1.6}$$

The association constant is related to the Gibbs free energy of binding

$$\Delta G_{bind} = -RT \ln K_a, \tag{1.7}$$

where R is the gas constant and T the absolute temperature. As seen from Eq. (1.3), K_a has units of concentration and therefore the value of ΔG_{bind} depends on the units used (i.e. the standard state).

The binding data are usually obtained by titrating together known quantities of reactants and measuring the concentration of either of the chemical species in the mixture. Various experimental techniques can be used for the concentration measurements, e.g. isothermal titration calorimetry, spectrophotometric methods, chemical shift in NMR and others.

The resulting data are commonly analysed graphically, illustrated in Fig. 1.1. The most natural displaying method is shown in Fig. 1.1 a,b, where the concentration of the bound ligand is plotted against the free ligand concentrations. The observed curved binding plots used to be difficult to analyze before the computer fitting programs. Scatchard therefore suggested a way to linearise the data by plotting the concentration of the bound ligand weighted by the concentration of the free ligand as a function of the
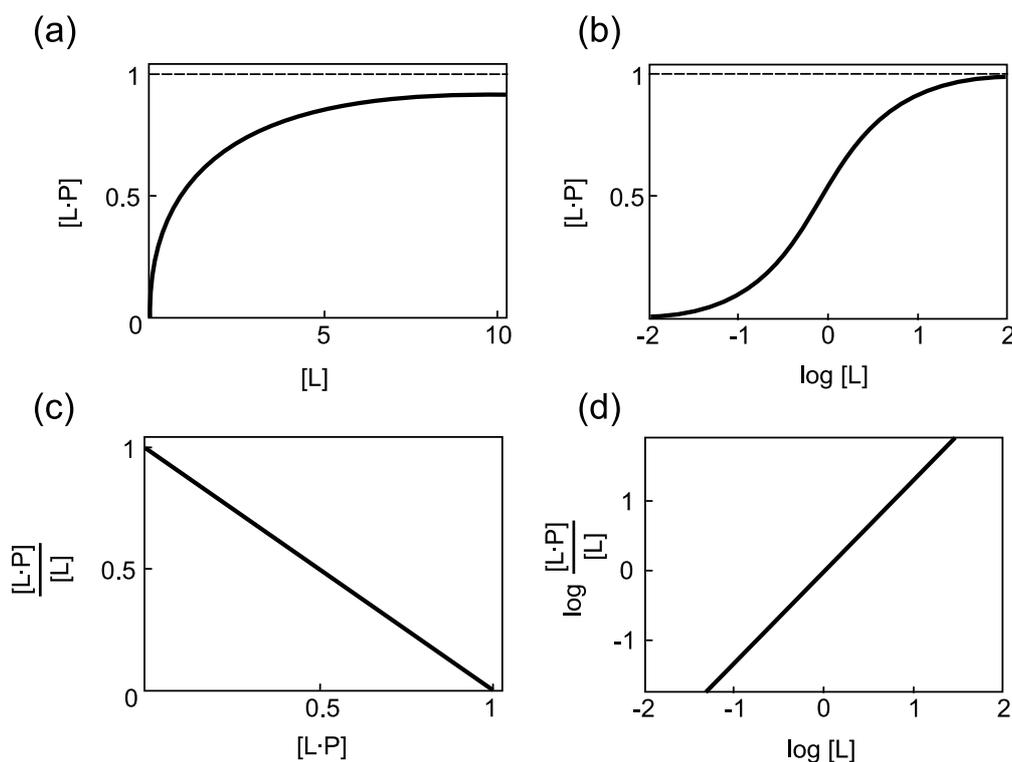
FIGURE 1.1: Examples of graphical methods used to analyse binding data. (a) The most natural way of displaying the data where the concentration of the bound ligand is plotted against the concentration of the free ligand. The usually observed hyperbolic relationship is shown. (b) A logarithmic scale is used for the free ligand concentration to display the whole binding range. (c) Scatchard plot. The negative slope yields the value of association constant and the intercept with the horizontal axis the value of maximal binding. (d) Hill plot used mainly for analyzing allosteric data (for details see Fig. 1.4). Here pictured for non-cooperative interaction. Adapted from [22].

bound ligand concentration; a method used until present time. The intercept of the Scatchard plot with the x-axis corresponds to the extent of maximal binding and the slope to negative association constant $(-K_a)$. When there is more than one binding site the Scatchard plot becomes nonlinear. Interpretation of the data in this case is explained later in this chapter (1.1.1). A particularly useful graph for analysing allosteric binding is the so called Hill plot (Fig. 1.1 d), because it provides a way to quantify the extent of cooperativity as we shall see later.

The mechanism of ligand binding to a protein was first dealt with by Fisher in 1894 [23]. He introduced a key-lock theory, which remained unquestioned for almost 60 years. This theory considers molecules as rigid structures. In order to bind a ligand the protein binding site must be a negative shape of the ligand so that the ligand fits into it like
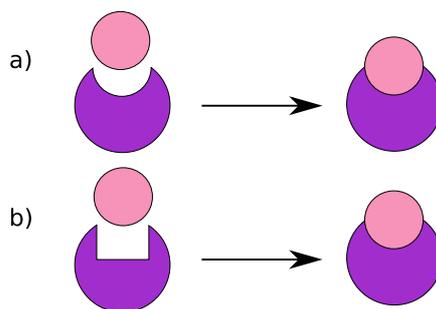
FIGURE 1.2: Schematic drawing of the key-lock (a) and induced fit model (b). (a) The ligand must be of exactly the negative shape of the protein binding site, (b) the binding site adapts its shape to better fit the ligand.

a key into a lock. The theory explained well the steric repulsion observed in organic reactions.

The key-lock theory was challenged by Koshland in 1958 [24] who introduced the concept of induced fit as an explanation of the specificity of the enzyme reactions. He came to the conclusion that a protein is more of a pliable rather than a rigid structure and the substrate binding may cause a substantial change in the three dimensional structure of the active site. A schematic illustration of the difference between the two models is shown in Fig. 1.2.

The ligand binding indeed commonly results in a modified 3d structure of the protein but it is now accepted that the process is lot more complicated still. Protein ligand association, like any other chemical reaction is a complex thermodynamic phenomenon with the equilibrium determined by the free energy. The Gibbs free energy is composed of an entropic and an enthalpic part

$$\Delta G = \Delta H - T\Delta S. \tag{1.8}$$

Changes in the structure described in the induced fit model contribute to the enthalpic part. More generally, we refer to a reaction as enthalpically favourable if $\Delta H < 0$, which corresponds to the lower internal energy of the product than of the reactant. An enthalpically favourable reaction simplistically involves creation or reinforcement of the interactions between the ligand and the protein and within the protein itself.

An entropically favourable reaction is such that $\Delta S > 0$. Entropy is a measure of the disorganization of the system and nature usually favours more disorganised systems.

Entropy can be calculated from the number of microstates available to the system

$$S = -k_B \sum_i P_i \ln P_i, \tag{1.9}$$

where index i runs over all accessible microstates and P_i is the probability of occupation of the particular microstate. Entropy is increased if the number of visited microstates increases. A ligand binding to a protein generally influences the thermal fluctuations of the protein, by either stiffening or softening the protein. The more commonly observed stiffening corresponds to reduced amplitude of fluctuations and thus implies a smaller number of significantly occupied microstates and reduced or unfavourable entropy. The entropic contributions are not accounted for in the induced fit model and are still occasionally disregarded in binding studies. They are, however, as equally relevant as enthalpic contributions, in particular in biochemical reactions involving macromolecules.

The enthalpy and entropy terms are intimately connected. An attractive interaction between two molecules is always, at least partly, compensated by an entropic cost of bringing the two molecules together. An entropic penalty is also paid for any loss of internal rotations and vibrations of the molecule and thus an enthalpically favourable reaction characterised by tight binding, that stabilises the protein structure is generally reflected in unfavourable entropy due to the loss of mobility.

The term enthalpy-entropy compensation is widely spread in literature but has to be regarded with caution. The effect is often observed as an artifact of wrong interpretation of data, in particular of the Arrhenius plots [25]. It can nevertheless play an important role in the reaction equilibrium and reliable methods, such as isothermal titration calorimetry, exist that can determine the thermodynamic parameters accurately. The methods most commonly employed to investigate details of ligand binding are reviewed in Sec. 1.4.

1.1.1 Multiple Binding Sites

Macromolecules often have more than one binding site and interact with multiple ligands. Association with multiple ligands can lead to a very complex behavior. In the simplest case the multiple sites are identical and have the same intrinsic affinities for each binding. There may also be distinct sites with different, and independent binding constants. But most intriguingly one ligand binding may influence the subsequent affinity for another

ligand. This cooperative effect is commonly referred to as allostery and constitutes the main focus of this thesis. Let us however begin our review with the simple case of independent binding sites.

Identical independent sites are common in proteins composed of several symmetrically arranged subunits. The observed binding curves are generally identical to the single ligand case, but the interpretation varies slightly. In particular, if we replace the concentration of the protein/ligand complex [L·P] in the Scatchard plot with the average number of ligand molecules bound per protein ν then the plot remains linear. The intercept with the vertical axis now however corresponds to nK_a and the intercept with the horizontal axis to n, where n is the number of binding sites and K_a is the intrinsic association constant (Fig. 1.3 a).

The association constant of multiple binding can be reported as an overall constant K_a or stepwise (intrinsic) constant of the j-th binding step K_a^j. The first ligand binding to a protein has n binding sites available and thus the association constant of the 1st step K_a^1 is

$$K_a^1 = \frac{nk_a}{k_d} = nK_a. \tag{1.10}$$

The n-th ligand has only one site available, but there are n ways a ligand can dissociate. Thus

$$K_a^n = \frac{k_a}{nk_d} = \frac{K_a}{n}. \tag{1.11}$$

When binding sites are independent but non-identical the Scatchard plots become more complex (Fig. 1.3 b). The stoichiometries and affinities of the bindings can be extrapolated, but the extrapolation has to be done cautiously. More complicated binding models in combination with multiple titration ITC experiments or NMR are usually used in order to obtain a complete picture.

A useful measure of relative affinity of two ligands is the Gibbs free energy

$$\Delta\Delta G = RT \ln \frac{K_a^{L1}}{K_a^{L2}}, \tag{1.12}$$

a quantity independent of the chosen units of the association constants.
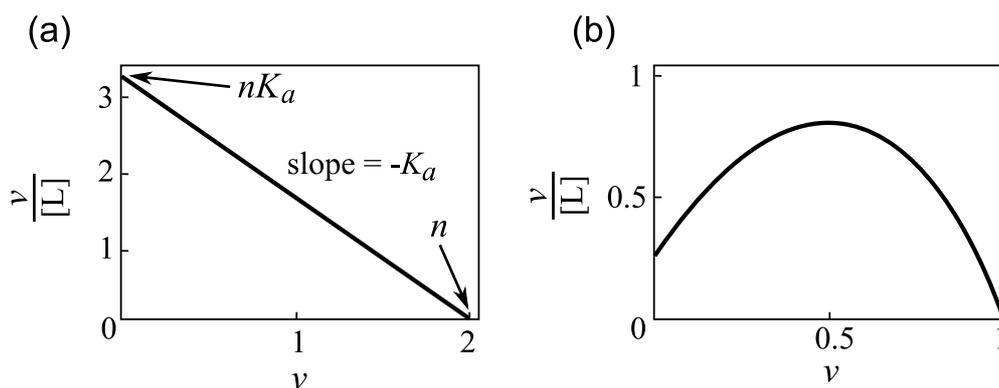
FIGURE 1.3: Scatchard plots of binding of a ligand L to a protein P, where ν is the average number of ligand molecules bound per protein and [L] is the concentration of free ligand. (a) Scatchard plot of two independent sites. The slope of the binding curve is $-K_a$, intercept with the horizontal axis is the number of ligands bound n and the intercept with the verical axis is nK_a. (b) Scatchard plot in the case of cooperative binding. For analysis of cooperative binding other methods such as the Hill plot are usually used. Adapted from [22].

1.2   Cooperative Binding

The affinity for a particular ligand can be changed by other ligands binding to the protein. This effect is known as cooperativity or more commonly allostery. The word allostery comes from greek *allos* ($\alpha\lambda\lambda o\varsigma$) meaning "other", and *stereos* ($\sigma\tau\epsilon\rho\epsilon o\varsigma$), "solid (object)" referring to the fact that the bindings occur at different sites of the protein (or other macromolecule). It constitutes a crucial mechanism of regulation in biological systems. Biological systems are complex networks where abundant external inputs have to be amplified or conversely nullified in order to maintain the organism's optimal performance. Allosteric effects provide a mean to tune individual steps and thus regulate the overall reaction pathways.

Binding of a ligand (so called effector) to a protein can result in increased or decreased affinity for the second ligand (so called substrate). The former is referred to as positive allostery (or cooperativity), the latter as negative allostery (or anti-cooperativity). The two ligands can be identical, the allostery is then said to be homotropic or non-identical resulting in heterotropic allostery.

The sign (positive or negative) and extent of allostery is defined by the allosteric free energy $\Delta\Delta G$ (Eq. (1.12)). For the homotropic allostery $\Delta\Delta G$ is alternatively defined as

the difference between the free energy cost of each binding

$$\Delta\Delta G = (G_2 - G_1) - (G_1 - G_0), \qquad (1.13)$$

where G_0 is the free energy of the free protein, G_1 of singly and G_2 of doubly bound. If the cost of the second binding is higher than of the first, i.e. if first binding inhibits the second the resulting allosteric free energy is positive and vice versa. We preserve this sign notation throughout the thesis

$$\text{if } \Delta\Delta G > 0 \Rightarrow \text{negative cooperativity} \qquad (1.14)$$

$$\text{if } \Delta\Delta G < 0 \Rightarrow \text{positive cooperativity} \qquad (1.15)$$

The absolute value of the allosteric free energy corresponds to the strength of the allosteric cooperativity. The larger values correspond to the larger influence of the first binding event on the second.

In the case of heterotropic allostery typically a different notation is adopted. A protein with an effector bound is called the holo-protein and the protein without the effector bound the apo-protein. The allosteric free energy is then the difference between the free energy cost of the holo-protein and apo-protein binding the substrate

$$\Delta\Delta G = \Delta G_{\text{holo}} - \Delta G_{\text{apo}}. \qquad (1.16)$$

The same sign convention applies. If the effector increases the affinity for the substrate then $\Delta\Delta G < 0$ and vice versa.

The history of allostery is intimately connected with discovering and explaining the oxygen binding properties of haemoglobin, the most famous allosteric system. Many models originally created for this system have been found to have implications for all allosteric systems and are widely used until today.

The first report of cooperative reaction was published in 1904 by Christian Bohr, father of atomic physicist Niels Bohr, who observed the association of oxygen with haemoglobin [26, 27]. Bohr carefully measured the dependence of fractional saturation of haemoglobin with oxygen on the oxygen pressure and observed a sigmoidal curve, in contrast with the usually observed hyperbola-like shape (Fig. 1.4). He reasoned that this shape must be a sign of cooperative binding, i.e. the affinity for the oxygen increases with the number

of oxygens already bound. In the same article Bohr reported that presence of carbon dioxide lowers the oxygen affinity.

The earliest attempt to explain the sigmoidal nature of the curve came from Archibald Hill in 1910 [28]. Hill assumed that haemoglobin binds n molecules in all-or-none fashion. The equilibrium is then described by

$$\mathrm{P} + n\mathrm{L} \rightleftharpoons \mathrm{P} \cdot \mathrm{L}_n \tag{1.17}$$

and the association constant

$$K_a = \frac{[\mathrm{P} \cdot \mathrm{L}_n]}{[\mathrm{P}][\mathrm{L}]^n}. \tag{1.18}$$

By combining Eq. (1.6) and (1.18) the saturation fraction of the protein molecule (e.g. haemoglobin) with the ligand (e.g. O_2) can be written as

$$Y = \frac{K_a[L]^n}{1 + K_a[L]^n} = \frac{[L]^n}{K_d + [L]^n}. \tag{1.19}$$

The more commonly used version with the dissociation constant (the right hand side) is known as the Hill equation. In order to plot the binding data in the form of the Hill plot we express the ratio on the vertical axis in terms of the saturation fraction

$$\frac{[\mathrm{P} \cdot \mathrm{L}_n]}{[\mathrm{P}]} = \frac{Y}{1 - Y} = K_a[L]^n. \tag{1.20}$$

Taking the logarithm results in

$$\log\left(\frac{Y}{1 - Y}\right) = \log K_a + n \log[\mathrm{L}]. \tag{1.21}$$

In the hypothetical case of infinite cooperativity, assumed by Hill, the Hill plot would be linear with slope equal n. The binding, however, can never be totally cooperative. But the factor n may be taken as non-integral parameter describing the extent of cooperativity. This parameter is known as the Hill coefficient and provides a quantitative measure of degree of cooperativity. The resulting Hill plot is composed of several segments (Fig. 1.4). The steepest part has slope equal to n and occurs in the middle of the free ligand concentration range. The outer segments have slope equal to unity and correspond to the first and last ligand binding. Negative cooperativity results in Hill plots with slopes of less than unity.
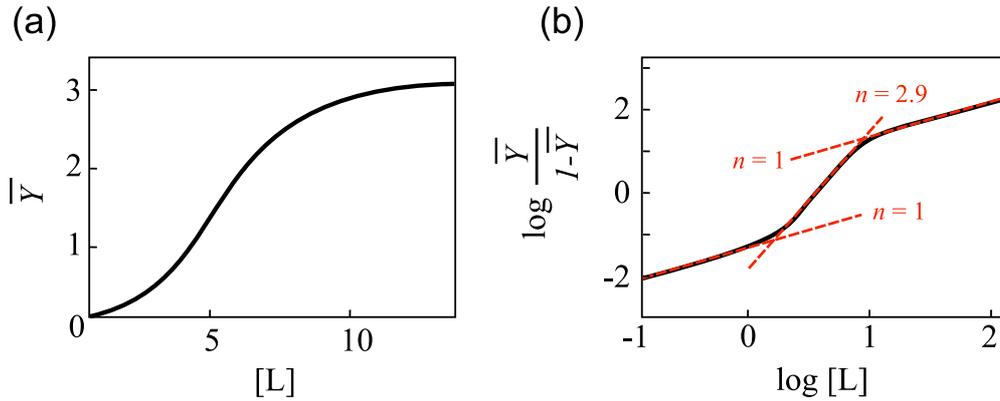
FIGURE 1.4: Binding curves for positively cooperative binding. (a) A typical sigmoidal curve observed when $\bar{Y}$, the average saturation fraction, is plotted as a function of ligand concentration. (b) Hill plot with a Hill coefficient $n = 2.9$, measured for human haemoglobin. Adapted from [22].

The next major contribution to understanding of the physiological properties of haemoglobin came from Gilbert Adair in 1925 [29]. He determined the molecular weight of haemoglobin and calculated that there are four binding sites for oxygen. He presented a model of allostery based on equilibrium constants that increase with the number of associated oxygen molecules. The Adair equation relates the degree of oxygen saturation Y and the partial pressure of oxygen p (used instead of concentration for gases)

$$Y = \frac{K_1 p + 3 K_1 K_2 p^2 + 3 K_1 K_2 K_3 p^3 + K_1 K_2 K_3 K_4 p^4}{1 + 4 K_1 p + 6 K_1 K_2 p^2 + 4 K_1 K_2 K_3 p^3 + K_1 K_2 K_3 K_4 p^4}, \tag{1.22}$$

where $K_1, \dots, K_4$ are the individual equilibrium constant for steps 1 to 4 respectively.

1.2.1 The Monod-Wyman-Changeux Model of Allostery

The intriguing problem of haemoglobin sparked interest in Jacques Monod, Jeffries Wyman and Jean-Pierre Changeux who suggested a model of allostery known as the two state concerted model, or simply as the MWC model [30]. They took inspiration from the induced fit ideas and refined them for the allosteric systems. Despite their simplified and, as later experiments revealed, imprecise assumptions about allosteric systems, their work became a cornerstone of modeling and understanding allostery. Due to its relative simplicity and direct interpretation of experimental data the MWC model

has enjoyed a wide spread popularity from its first publication in 1965 [30] until present times.

They assumed that allosteric systems are composed of two or more identical subunits with at least one axis of symmetry. Each subunit can exist in two (or more) conformational states, which have different affinity for the ligand binding and different quaternary structure (i.e. the intersubunit bonds). The states are denoted as T (tense) and R (relaxed) where R has higher affinity for the ligand. The two states with j ligands bound are denoted R_j and T_j. The main postulate is that the symmetry of the system has to be conserved, that is, all subunits have to be in the same state at all times. The switch between the conformations therefore must be concerted, hence the name of the model. In the absence of the ligand, the two states T_0 and R_0 exist in equilibrium defined by the equilibrium constant L. The T_0 state with low affinity for the ligand is favoured ($L < 1$). The affinity of the R state for the ligand is higher than of the T state by factor c ($c > 1$). The equilibrium of the states R_1 and T_1 is therefore shifted towards R_1 and is determined by the constant Lc. The unligated sites of the R_1 state therefore have increased affinity for the ligand. The full scheme of this model is shown in Fig. 1.5.

Importantly, L and c are the only two parameters of the model. They can be extracted relatively easily from the experiments and fully characterise the allosteric system.

The model was primarily aimed at explaining homotropic allostery. The authors suggest that heterotropic ligands can use the same mechanism by stabilizing either the T or the R state to produce positive or negative cooperativity.

In 1965 when this model was published it was a major success. Allostery had been only observed for symmetric oligomers and even the fact that the model did not account for negative homotropic cooperativity (a ligand inhibiting binding of a second identical ligand) did not matter, because no example of such behaviour was known at that time. Later observations however challenged these and other limitations and slowly forced a new view on the mechanism of allostery.

1.2.2 The Koshland-Nemethy-Filmer Model of Allostery

Shortly after the publication of the MWC model Koshland, Nemethy and Filmer proposed an alternative explanation of allostery, usually referred to as the KNF or the sequential model [31]. It has been created for allosteric proteins composed of two or

(a) The MWC concerted model

(b) The KNF sequential model
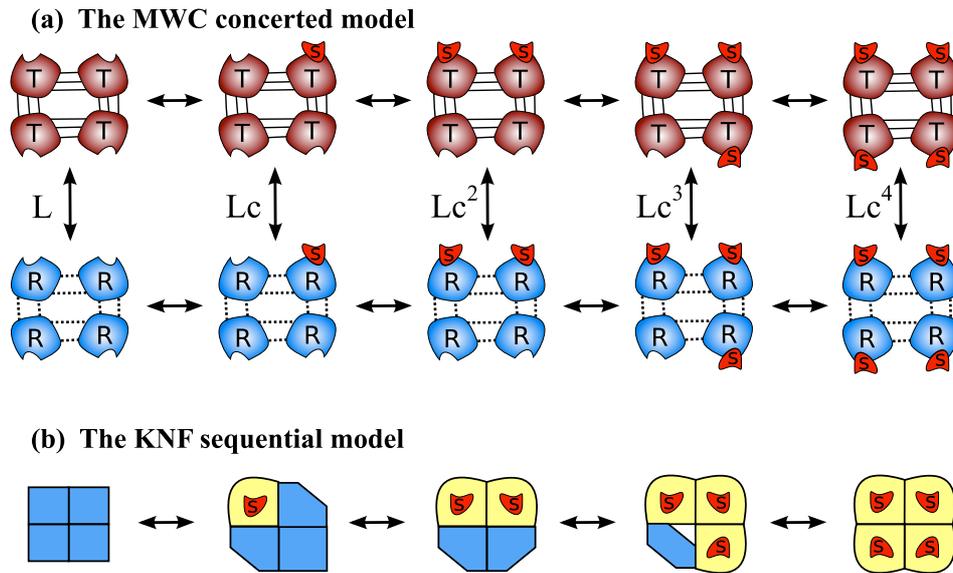
FIGURE 1.5: Schematic drawing of the MWC and KNF models. (a) The conserved symmetry of the protein within the MWC model requires all subunits being always in the same state, either in the T state (brown) or R state (blue). The R state is characterised by higher affinity for the ligand. (b) The KNF model: each ligand induces a conformational change in the subunit to which it is bound and in the neighboring subunits. The affinity for the ligand thus increases sequentially. The geometry is defined by allowed interactions rather than by the protein structure.

more subunits. It assumes that binding of a ligand can directly influence the tertiary structure of the neighbouring subunit and thus the affinity of this subunit for the ligand. The neighbour is defined by the allowed interactions between the subunits rather than by their position in space.

The model predicts the average number of molecules of substrate bound per protein N_S as a function of ligand concentration and four parameters: the equilibrium constant K_S for binding of the substrate S to the individual subunit, the equilibrium constant for the conformational change K_t and interaction constants K_{AB} and K_{BB} between the conformations A and B. The exact expression for N_S however depends on the arrangement of the subunits within the protein, the original article gives formulae for tetrahedral, square and linear arrangement of 4 subunits. The propagation of conformational changes in the square arrangement of subunits is illustrated in Fig. 1.5.

The MWC and KNF models are both based on the idea of conformational change. The conformational change in the KNF model, occurs only as a consequence of ligand binding and extends directly to the other subunits. The extent of conformational change is thus

proportional to the number of ligands bound. The MWC model is, on the other hand, founded on the idea of two symmetric quaternary structures, present in equilibrium even in the absence of a ligand. The conformational change from the low affinity T-state to the high affinity R-state then tends to take place when Lc^i becomes greater than unity, which can in principle occur after any binding step.

Clearly, the KNF model is mathematically more complicated, the equation for N_S depends on the interaction model and 4 parameters, but less restrictive than the MWC model. The restrictions that the MWC model imposes are advantageous in that only two parameters need to be fitted, but the model does not account for negative cooperativity, cannot explain changes in tertiary structure and preferential binding to a specific subunit, all of which have been observed in numerous proteins.

Even though one can argue the KNF model is better as it explains wider range of effects, thanks to its mathematical complexity it has never achieved as much success as the MWC model.

1.2.3 Current View Of Allostery

The previous two models represent classical examples of the static view of allostery. They assume that conformational change provides the only means of communication between the two binding sites. This change in the three dimensional structure of the protein is triggered by the first binding, or in the case of more than two binding sites by any subsequent binding. Thermodynamically this corresponds to a purely enthalpic effect. However, as we explained before, the entropic contributions may play just as important role in the ligand binding. This observation was first made by Weber in 1972 [2]. He pointed out that every protein constantly fluctuates between a number of conformations and we only measure the average one. He argued that binding of a ligand can stabilise a certain conformation and thus aid binding of another molecule.

A different approach was taken by Cooper and Dryden who in 1984 introduced the notion of dynamic allostery and evaluated its potential importance based on basic statistical thermodynamics principles of ligand binding, very similar to the ones we use in this work. They focused on binding of identical ligands to distant sites for which the allosteric free

energy can be expressed as

$$\Delta\Delta G = -k_B T \ln\left(\frac{Z_0 Z_2}{Z_1^2}\right), \qquad (1.23)$$

where Z_i is a partition function of state i; $i = 0$ refers to free, $i = 1$ to singly and $i = 2$ to doubly liganded protein.

They argued that of the number of physical phenomena that may contribute to the allosteric effect, only the changes in vibrations and conformational states of the protein result in significant allosteric free energy. Contributions from more subtle changes in the dynamics that may occur due to redistribution of conformational substates can be important too, but arise through similar mechanisms as the vibrational contribution. Let us review here only the outcome of their calculations on the previously neglected changes in vibrational spectra.

The internal motion of a protein can be decomposed into a spectrum of vibrational normal modes. The classical partition function of a single thermally excited normal mode with frequency ν is

$$Z(\nu) = \frac{k_B T}{h\nu}, \qquad (1.24)$$

where h is the Planck constant. If this low frequency normal mode undergoes a frequency shift $\nu_0 \to \nu_1 \to \nu_2$ as a result of the sequential ligand binding the allosteric free energy becomes

$$\Delta\Delta G_{\text{vib}} = -k_B T \ln\left(\frac{\nu_1^2}{\nu_2 \nu_0}\right). \qquad (1.25)$$

The fraction on the right hand side can differ from unity, resulting in finite allosteric free energy. This demonstrates that an alternative communication pathway between the binding sites exists and conformational change is not required for allosteric effects.

The result also provides a condition for the sign of allostery; if $\nu_1^2 > \nu_2 \nu_0$ then the allosteric effect is positive and vice versa. Additionally, the equation can be used to estimate the extent of this effect in proteins. Assuming 10% increase in frequency upon each binding, the allosteric free energy per mode equals $\sim 0.01\, k_B T$. In a average protein, there are hundreds of classical normal modes and the effect can thus give rise to biologically relevant free energies of several $k_B T$.

This insightful article has however not resulted in a radical change in attitude towards allosteric effects. It instead took many years till their demonstration had been seriously
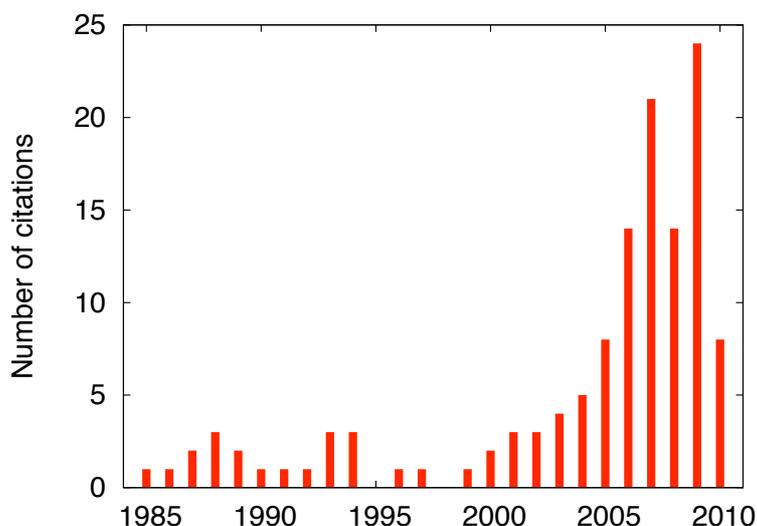
FIGURE 1.6: Increase in annual number of citations of the article "Allostery without conformational change" by Cooper and Dryden [18].

acknowledged by at least a part of the scientific community (Fig. 1.6). This delay was partly due to the lack of experimental techniques able to verify their hypothesis. It took a decade for experimental techniques to advance enough to be able to reliably observe that structural fluctuations can become modified upon ligand binding [32] and even longer to connect this phenomenon with allostery [4].

The new view of allostery that emerged from the novel experimental techniques was summarized in 1999 by Freire in the so called dynamic population shift model [14]. In his article he describes a protein as a statistical system, undergoing permanent fluctuations. The fluctuations can be regarded as local unfolding events and their probability is given by the Boltzmann factor. A change in the protein conditions, such as binding of a ligand is reflected in the most probable distribution of the states. The redistribution of the states can affect also regions distant from the ligand binding site. Thus the "pre-existing equilibrium" favouring binding of a substrate can be stabilised. The communication proceeds only via a few residues, the so called cooperative pathway, connecting the two binding sites. Although his ideas were not completely original they were supported by much experimental [33–35] and computational evidence [36, 37] and therefore meant a substantial departure from the traditional understanding of ligand binding as an induced fit event.

Molecular flexibility has become accepted as crucial to allostery but only as providing a route to conformational change. In the view of the population shift model, flexibility gives rise to population of conformational states between which the protein fluctuates. The rates of fluctuations depend on the binding state of the protein and a ligand binding can stabilize certain conformation. But as Cooper and Dryden pointed out this is not the only role of flexibility. Dynamics can contribute to allostery even in absence of observable change in average conformation. In the early 2000s it became increasingly more apparent that signals can indeed be transmitted through proteins by changes in dynamics [15, 38, 39]. If a last piece of evidence was needed it was provided by Popovych *et al.* [4]. They presented a complete study of an allosteric protein (CAP) that performs its cooperative binding without a substantial conformational change. This means that an alternative mechanism of communication between the sites is taking place and it can only be provided by altered dynamics. In other words the allostery is entropy driven in the way described by Cooper and Dryden. We tried to gain a deeper understanding of the mechanism of allostery in CAP by using a coarse-grained model; the results of our calculations along with details of the experiments are presented in Chap. 2.

The example of the CAP protein is an extreme case where structural change plays no part in the propagation of the signal. However, dynamics and structure are of comparable importance to the allostery and a majority of proteins uses both means of communication to a certain degree. Tsai *et al.* suggested a classification scheme where they divided allosteric proteins into three classes (Type I, II and III) based on where is their allostery located on the spectrum with the extremes of enthalpic and entropic effect [40]. Type I is governed by entropy, i.e. there is no or only subtle conformational change observed, Type II is driven by both entropy and enthalpy and Type III is predominantly driven by enthalpy.

Cooper and Dryden formulated the general principles of allostery propagation through modified dynamics but did not apply their formalism on any specific systems. Hawkins and McLeish have expanded their ideas into the form of concrete models of two classes of proteins: rigid dimers with application to the lac repressor [38] and coiled coils [41]. They also elaborated on the role of fast, localized motions and suggested how these may become coupled to the slow global motions. They concluded that the resulting effect can be amplification of the entropic allostery and can give rise to compensating entropic and enthalpic terms. They exemplified the model on the DNA binding met repressor protein [42].

Another view of allostery that has gained significant popularity is based on the idea that the two binding sites are linked by a set of residues, the so-called allosteric pathway, along which signals are communicated [7, 16, 17]. Communication can proceed via structural changes and/or altered vibrations of the linked residue. The networks commonly consist of evolutionary conserved residues and are very sensitive to single mutations rendering them popular amongst biologists and computational scientists alike. However, the majority of these studies concentrate on detecting the pathways rather than the mechanism by which the interaction proceeds.

Not only has the understanding of the mechanism of cooperativity progressed, but so has the perception of an allosteric system. The traditional view of an allosteric protein as oligomer with multiple identical domains has been overthrown by Volkman *et al.* in 2001 [35] who proved the existence of allostery in a single domain protein. Gunasekaran *et al.* [39] went a step further and suggested that allostery may be an intrinsic property of all non-fibrous proteins. In the opinion of Gunasekaran *et al.*, using appropriate ligands, point mutations or external conditions may result in allosteric behaviour in any protein. Such a hypothesis has large potential implications for pharmaceutical industry and the shift in focus in new drug development toward allosteric mechanisms has already been noticeable [17, 43, 44].

We have explained why both the enthalpic and the entropic parts of the free energy play a key role in allosteric effects. This has now become widely accepted, but the mechanism by which is the communication established remains elusive. It is precisely those systems, where entropy plays a major role, that are not well understood and therefore offer a great opportunity for substantial advances in understanding.

1.3 Protein Dynamics

It has been nearly 200 years since the botanist Robert Brown systematically described thermal motion of pollen granules and over 100 years since Albert Einstein explained it. Yet it had taken a long time to accept that thermal motion plays an essential role at the intramolecular level in biochemistry. Weber in 1972 was possibly first to point out that macromolecules undergo permanent fluctuations that can be essential for their function [2]. In 1975 Austin *et al.* supported the hypothesis by showing that an ensemble

of protein molecules in the same conformational states is not homogeneous, i.e. some molecules occupy different conformational substates [45].

The potential significance of protein dynamics was occasionally mentioned in literature during the 70s and 80s, but attention then was focused on determining the macromolecular structure. This was partly due to the absence of an experimental technique able to observe protein dynamics directly. X-ray crystallography can identify flexible regions but the results are rarely conclusive. A revitalisation came in the form of the discovery of a method able to interpret NMR data in terms of protein dynamics.

Lipari and Szabo [46, 47] noticed that NMR may be the right method for the protein dynamics investigation because it is very sensitive to protein motions, can resolve independently the dynamics of different residues on the protein because of their chemical shifts, and can be carried out in a solution and at room temperature. They developed their ideas into a practical tool, and in a few years NMR has become the most popular method for investigating protein dynamics. The revelation of the functional importance of flexible regions with NMR had a large impact on modern protein science, as has been stressed, e.g., in recent special thematic issues of the journal *Science*: (1) Biochemistry: Tools for New Frontiers published on 14 April 2006 and (2) Protein Dynamics released on 10 April 2009.

The less intuitive and less well understood entropic part of the free energy of protein-ligand association comprises terms associated with the protein, the ligand and the solvent. The latter two can be seen as the entropic cost related to the transformation of the ligand from the free to the structured (bound) state and the accompanying solvent release from the binding site. These may contribute significantly to the free energy [48], but are of secondary significance in the studies of allostery, in particular homotropic allostery, because the entropic cost is nearly identical for the two binding events. The least understood part of the entropic term is associated with the protein and is usually referred to as the conformational entropy. It involves the entropic cost of changes in the protein internal dynamics induced by the ligand binding. Here we focus on the processes occurring in the macromolecule that lead to modification of its dynamical behavior.

Macromolecular dynamics is a very vague and often misused term. When referring to dynamics we choose to follow definition introduced by Henzler-Wildman and Kern, that "protein dynamics is any time-dependent change in atomic coordinates" [49]. This definition includes both, equilibrium fluctuations and non-equilibrium phenomena. Here

(a)

(b)
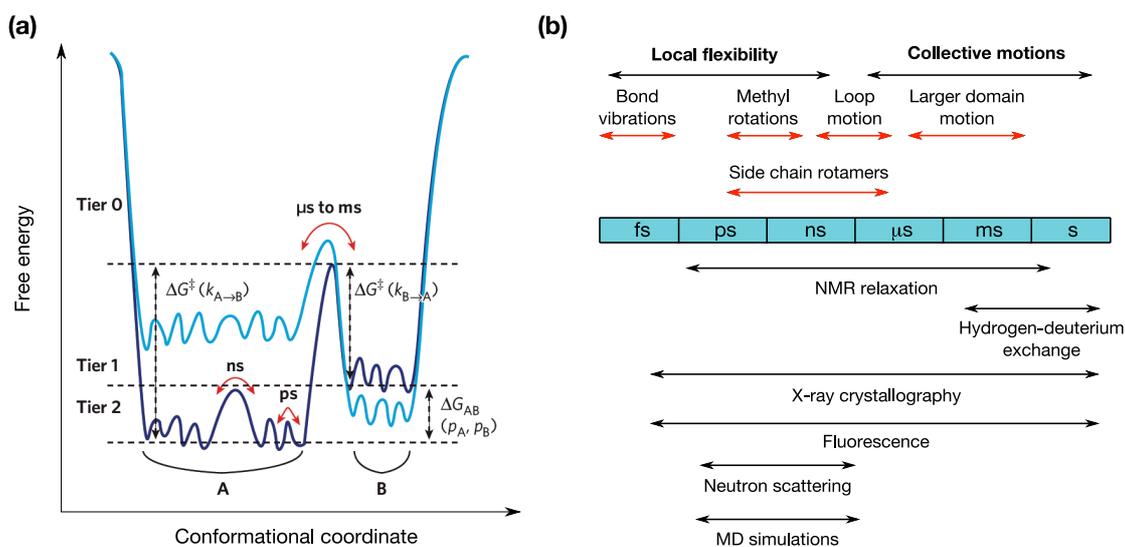
FIGURE 1.7: Illustration of protein dynamics timescales. (a) One dimensional crossection through a free energy landscape of a protein shows the classification of local minima into three tiers. Motion within tiers 1 and 2 is referred to as fast dynamics and motion between different wells of tier 0 are denoted slow. (b) A time line describing protein dynamics and experimental techniques that can access motions on respective timescale. Adapted from [49].

we however only discuss the equilibrium fluctuations since they appear to govern the majority of biological processes including allosteric effects [49].

The equilibrium fluctuations are conventionally divided into slow and fast. Let us adhere to the formalism of Henzler-Wildman and Kern who define these terms based on the depth of the local minima of the free energy landscape in which the fluctuations occur (Fig. 1.7).

Slow fluctuations can be seen as conformational transitions between tier-0 wells which are separated by free energy barriers of several $k_B T$. They occur at μs-s timescales and are considered biologically very relevant because many functional processes such as protein folding, protein docking and allosteric transitions take place on similar timescales. The motions on these timescales are collective, involving the whole protein domains.

Fast fluctuations are defined as fluctuations within the tier-0 wells. These motions are characterised by smaller amplitude and ps-ns timescales. Physically they involve motions of small groups of atoms such as side chains or loops. We note that even faster fluctuations are present in proteins, such as atomic bond stretching that occurs

on femtosecond timescales. These motions are however not thermally excited and thus lie beyond the interest of this thesis.

In this thesis, we are primarily interested in collective motions occurring at ns-μs timescales but also in faster motions such as the side chain vibrations and loop motions happening with ps-ns frequency. The collective motions are particularly important since their perturbation is felt at large distances, in contrast with fast modes which are localised (involve only a few atoms) and thus unlikely to transmit the allosteric signal across large distances. Daniel *et al.* [50] reviews several experimental results and concludes that motions on timescales shorter than 100 ps are not required for enzyme function. Which motions are active in the allosteric signaling is disputed [4, 35, 51, 52], but the most likely answer is that to a certain extent motions from all parts of the spectrum play a certain role.

The motion of individual atoms or even concerted motion of larger biomolecular domains still cannot be observed directly. A number of sophisticated experimental techniques has however been developed able to measure certain physical properties as a function of time from which the dynamics can be inferred. We introduce some of these methods along with most important and influential experiments in the following section.

Computational simulations provide a prominent set of techniques capable of observing and studying protein motion directly. They open the window into the atomistic world of the molecules and allow for the observation of the macromolecular machines at work. Some of the techniques developed to perform and analyze the computer simulations are reviewed in the next chapter.

1.4   Experimental Techniques for Protein Dynamics and Allostery Investigation

The number of experimental techniques able to probe protein dynamics has grown rapidly over the past 20 years and it is beyond the scope of this thesis to review them all. Let us therefore focus on the major ones used to assess the role of dynamics in ligand binding.

The two most common techniques are isothermal titration calorimetry (ITC) and nuclear magnetic resonance (NMR). ITC is probably the only techniques where the free energy

(ΔG), enthalpy (ΔH) and entropy (ΔS) of the reaction are measured in one experiment, rendering it very useful for allostery consideration. However, the entropic term however includes the contributions associated with ligand and desolvation of the binding pocket and is thus not a direct measurement of the protein dynamics. NMR, on the other hand, is in principle able to measure protein dynamics directly. The particular techniques are however substantially complicated and the interpretation debated.

1.4.1 Nuclear Magnetic Resonance (NMR)

The major advantage of NMR is that the experiment is conducted in solution at room temperature and many different techniques can be employed to access a wide range of timescales. Some methods are well established but many have been developed only recently with the aim to both, widen the range of timescales measured and improve the quality of the data collected. Most biological processes, such as ligand binding, allostery, protein-protein interactions or folding happen on μs-s timescales. The techniques accessing these and slightly faster backbone motions are crucial in addressing the research questions of this thesis. However fast, side chain, motions may in many cases be tightly correlated with the slow motions [42] and therefore are of high interest to us too.

Here we briefly introduce the physical principles of NMR and point out a few techniques that specialize in dynamics measurements. We also review several applications illustrating the scope of use of these techniques. For the purposes of this thesis, we will treat the review of NMR techniques at a classical level. While a quantitative and accurate implementation of the techniques requires a quantum-mechanical analysis, the corresponding classical picture serves to illustrate the abilities and scope.

A protein or any other molecule is a system of atoms each of which have a nucleus with intrinsic spin $\mathbf{S}$ and magnetic moment μ. In equilibrium the nuclear magnetic moments point in all possible directions; the distribution of the directions is isotropic. When placed in external magnetic field $\mathbf{B_0}$, the nuclear magnetic moments interact with the field and start to precess around the field with a characteristic frequency called the Larmor frequency (Fig. 1.8). Furthermore the entire sample acquires a small net nuclear magnetic moment along the field. When the magnetic field is switched off the net nuclear magnetic moment relaxes. The relaxation of each nucleus depends on the fluctuations and distance of the neighboring nuclei and the overall tumbling of the molecule. This relaxation time, called the longitudinal relaxation time or T_1, is measured
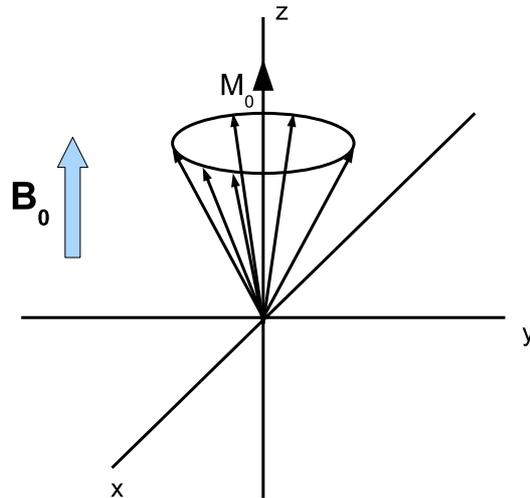
FIGURE 1.8: The precession cone of nuclear spins in the external magnetic field $\mathbf{B_0}$. The distribution of the spins in x and y direction is isotropic. The external magnetic field causes an alignment of the spins in the z direction resulting in the net magnetic moment of the sample M_0. The relaxation of M_0 is referred to as the longitudinal relaxation.

and in combination with the other relaxation parameters interpreted in terms of protein structure and protein dynamics.

In order to gain more detailed information about the sample, additional parameters sensitive to different physical processes are measured. First is the transverse relaxation or T_2, which describes the relaxation in direction orthogonal to the external field. This relaxation process is rather complex consisting of gradual loss of coherence in the phase of spin precession. Since this only occurs due to interaction of magnetic fields of different spins it is commonly referred to as the spin-spin relaxation time. In contrast, the longitudinal relaxation occurs due to energy exchange with its surroundings, the lattice, and is thus often referred to as the spin-lattice relaxation.

A third parameter, called the Nuclear Overhauser Effect (NOE) completes the set of parameters measured in a typical NMR experiment. It consists of applying a weak radio frequency field at the Larmor frequency of one of the spins of the system. The resonance has a strong effect on the spatially proximate spins and thus the measurement yields information on the local distances over time. In some specific applications other parameters are measured. These and the traditionally measured parameters and the techniques used to interpret them are explained in much greater detail elsewhere, e.g. [53].

A number of physical factors influence the relaxation: (1) dipole-dipole interactions strongly dependent on the distance between the nuclei, (2) chemical shift anisotropy, a result of shielding of a nuclei by the neighboring molecules, (3) quadrupolar interactions that influence only nuclei with spin ≥ 1 and (4) conformational exchange.

Evaluating individual contributions from the above factors to the overall relaxation provides the sought-after information on structure and dynamics of the molecule. However, it is a complicated task. In order to facilitate the decomposition of the NMR signal specialists have developed a string of techniques, generally involving various radio frequency pulses.

The methods are conventionally divided into three categories according to the timescales they are able to observe: fast dynamics (motions on ps-ns timescales), slow dynamics (μs-ms timescales) and very slow dynamics ($\gg$ms) that is generally referred to as the hydrogen-deuterium exchange. For a recent review of the methods see [54].

The most commonly used methods to investigate fast backbone motions are the reduced spectral density mapping (RSDM) [55] and "model free method" of Lipari and Szabo [46, 47]. These two methods measure dynamics of ^{15}N by evaluating the relaxation times T_i for the spin pair ^{1}H- ^{15}N. The model free formalism is preferred over RSDM because it provides information on the nature of the motion in addition to the timescale. Recently the field has been expanded by new techniques that yield complementary information to the ^{15}N measurements, e.g. techniques measuring carbonyl-C_α bonds dynamics [56] or correlated motions of successive protein residues [57].

The side chain dynamics can be probed by several labeling techniques [58]. They measure primarily the quadrupolar interactions of deuterium, used as a label.

The methods investigating the μs-ms (slow) dynamics make use of the large amplitudes of the fluctuations causing the chemical shift to be distinct in each conformation. They are collectively referred to as the relaxation dispersion measurements. By applying a cleverly constructed sequence of radio frequency pulses the difference between the shift becomes amplified and can be easily detected. The most common sequence is the Carr-Purcell-Meiboom-Gill (CPMG) sequence. Other sequences can be used to improve the quality of the data [59].

The CPMG sequence yielded the key results for example in the experiments of Eisenmesser *et al.* who showed the importance of slow dynamics in catalytical function of the

protein CypA [9]. Other groups have measured both slow and fast dynamics to characterize the role of dynamics in ligand binding. Allostery-related examples include a study of Malmendal *et al.* of calmodulin cooperatively binding calcium [33]. They showed that conformational change is accompanied by a significant reduction in the backbone motions on the μs-ms time scale upon Ca^{2+} binding. In another study Volkman [35] measured slow and fast motions of the signaling protein NtrC and observed that fast motions remain nearly unchanged but the slow motions get suppressed upon NtrC activation. Other studies showed that binding can be accompanied by modification of fast dynamics only [8, 51]. The majority of these methods were employed to address the change of dynamics during the allosteric event of a small cAMP molecule binding to the CAP protein [4]. This system constitutes a main example in Chap. 3 and the experimental results are thus reviewed in greater detail there.

One way to observe the very slow dynamics exemplified by the conformational transitions of a protein is to dissolve the protein in a deuterated buffer. The exchange rates between hydrogen and deuterium then depend on the amide exposure to the solvent. Several experimental methods are available to measure the exchange rates, such as mass spectroscopy but the most accurate is the site-specific NMR. Hydrogen-deuterium exchange rates have been measured with NMR for multiple proteins and yielded a high resolution dynamic measurements across a wide range of timescales (e.g. [4, 60]).

From our point of view, however, all NMR methods have a common drawback; they do not provide information on modal dynamics and instead report on local, spatial dynamics. Typically the result of an NMR dynamics experiment is a relative measure of mobility of individual residues within a certain frequency range (see e.g. the results of NMR experiments on the CAP protein performed by the group of Kalodimos and reproduced in Figs. 3.8, 3.9).

A first attempt to provide this, for most purposes, more meaningful description in terms of collective modes and uncorrelated local motions within a protein was recently reported by the group of Emsley [61]. They argued that solid state, or more precisely microcrystalline NMR is better suited for determination of the modal dynamics than the traditional solution experiments because the overall tumbling of the molecule is missing. They showed that small amplitude collective motions contribute to the solid state NMR relaxation rates and are detectable. They use the global modes obtained from computer simulations and show that they can explain a large part of the relaxation rates determined. They however have not yet developed a method that unambiguously

determines the nature of the motions, partly due to the lack of observables in the solid state NMR [62].

1.4.2 Isothermal Titration Calorimetry (ITC)

The starting point for analysing an allosteric system is the knowledge of the thermodynamical parameters of the reactions. These are most reliably provided by the technique of isothermal titration calorimetry (ITC). In a single experiment ITC measures the free energy, entropy and enthalpy of the reaction. ITC is also well suited for determining stoichiometry of the reaction, rendering it ideal for observing systems involving multiple binding events, such as allosteric systems. An ITC measurement is performed under physiological conditions and the reactants do not require any chemical modification.

In a typical ITC experiment small doses of a ligand solution ($\sim 10\,\mu$l) are regularly added into a pool of dissolved protein (~ 10 ml). Upon adding the ligand, binding occurs, and a certain amount of heat q is released or absorbed, depending on the nature of reaction. The heat released is related [63] to the enthalpy of binding ΔH by

$$q = V\Delta H\Delta L, \tag{1.26}$$

where V is volume of the reaction cell, ΔL is the increase of concentration of bound ligand upon adding of the dose of ligand. The amount of free protein decreases upon each successive injection and so does the amount of heat absorbed or released. The produced heat is non-zero even after saturation due to dilution and mechanical effect. This residual heat needs to be subtracted from the previous results.

In order to determine the value of ΔL and consequently the correct thermodynamic parameters, an appropriate binding model has to be selected. This model relates ΔL to the association constant of the ligand, which in turn gives the value of ΔG. The binding model depends on number of binding sites per macromolecule and the presence of cooperativity. For an overview of binding models for single and multiple binding event see [64].

Analysis of the data results in value for ΔG and ΔH; the entropy change can be calculated from $\Delta G = \Delta H - T\Delta S$.

By repeating the experiment at different temperatures the change in heat capacity can be obtained too

$$\Delta C_p = \frac{\partial \Delta H}{\partial T} \tag{1.27}$$

ITC has been successfully used to study cooperative binding of many systems [64]. Examples include study of entropically driven negative allostery upon binding two zinc molecules Zn^{2+} to CzrA homodimer by Grossoehme and Giedrock [65] or enthalpically cooperative inhibition of assembly of a ternary complex of ferredoxin $NADP^+$ reductase, $NADP^+$ and ferrodoxin electron donor protein by $NADP^+$ [66].

It may be advantageous to combine ITC with NMR as has been shown by Tochtrop *et al.* [67]. The authors investigated an extremely cooperative binding of glycocholate to human ileal bile acid. The calorimetry data were unable to define the enthalpy accurately. However, in combination with NMR, the binding affinities could be determined for each binding site separately with high accuracy. Popovych *et al.* also used both techniques to analyze the CAP-cAMP system [4].

1.4.3   X-ray Crystallography

X-ray crystal diffraction is predominantly used for macromolecular structure determination. Over 85% of currently known protein structures have been determined with this method [68]. Along with the structure the diffraction pattern carries information on the protein dynamics, albeit in the crystalline state. It can be extracted by measuring the so-called B-factors or related Debye-Waller factors.

The main difficulty of a crystallographic experiment is that a protein has to be crystallized. It can take months to grow a sufficiently large crystal (~ 0.5 mm) [69]. An X-ray beam is then scattered from the crystal and the diffraction pattern recorded. The diffraction pattern carries information on the structure factors and the electron density of the molecule. The atom positions are fitted into the reconstructed electron density map. A resolution in Ångström is given for each structure determined with X-ray crystallography. A resolution of 2 Å means that two objects that are at least 2 Å apart can be distinguished and in practice this means that different residue side chains can be resolved. In order to distinguish individual atoms a resolution of 1 Å is required but such resolution is achieved only very rarely.

Flexible regions of a macromolecule are best indicated by "missing" electron density. Sometimes a region of the map has very low and even electron density meaning that atom positions cannot be resolved. This is usually explained by a high flexibility of the part of the molecule resulting in no average structure.

Atoms undergo thermal motions even within crystals and X-ray crystallography provides a method to measure their mean square displacement $\langle x^2 \rangle$. The values of $\langle x^2 \rangle$ are commonly reported together with the structural data in the form of B-factors sometimes also called the Debye-Waller factor. The B-factor is directly proportional to the mean square displacement and is defined as $B = 8\pi^2 \langle x_i^2 \rangle$ [70]. The B-factors however have at least four contributing factors, thermal vibrations, conformational variation, the defects in crystal lattice and errors in crystallographic phases. To separate the contributions and draw conclusions about the flexibility of the protein at physiological temperature is very difficult. We should therefore consider B-factors only with caution and try to employ other techniques for the determination of dynamics.

The B-factors have nevertheless found their application. For example, Wooll and coworkers used the B factors to point out that dynamics may be responsible for the allosteric mechanism of mammalian pyruvate kinase [71]. Hawkins and McLeish used B factors to parameterise their coarse-grained model of the lac repressor [38]. Other examples may be found in literature, although lately NMR data have been preferred.

1.4.4   Neutron Scattering

X-rays scatter from atoms proportionally to the density of electrons in the cloud surrounding the nuclei. This leads to poor resolution of atoms with low atomic number. Neutrons are, on the other hand, scattered by atomic nuclei and the scattering does not increase with the atomic number. Neutron scattering therefore offers an alternative route to structure determination, but is complicated by technical difficulties such as the need of a neutron reactor or neutron source and long exposure of the crystals to the neutron beams.

Flexibility of the proteins can be measured with neutrons too. Neutron scattering is in principle able to measure three types of scattering [72]:

1. Elastic scattering which reports on atomic motions with Ångstöm amplitudes and timescales of $10^{-12} - 10^{-9}$ s;

2. Quasielastic scattering, from which correlation times of diffusion motions can be calculated;

3. Inelastic scattering, reflecting fast vibrational modes.

Typically the elastic experiment is performed because it is most reliable and easiest to interpret. The mean square displacement $\langle x^2 \rangle$ is calculated from the angular dependency of the scattered elastic incoherent intensity [73]. $\langle x^2 \rangle$ is then recorded as a function of temperature. The slope of the curve is used to calculate the effective mean force constant $\langle k' \rangle$, defining mean molecular rigidity.

Neutron scattering is also used to study dynamical transitions. It has been observed that proteins undergo an abrupt change in dynamical behaviour at a certain low temperature, generally in the range of 180-220 K [50]. Below the transition temperature atoms are considered to vibrate purely harmonically while above anharmonic motions occur. This is sometimes used as evidence that anharmonic motions are required for protein function [50].

Inelastic neutron scattering has been used to demonstrate that dynamics of a protein can vary upon ligand binding [74]. Small angle neutron scattering was used by Consler *et al.* to study the radius of gyration of pyruvate kinase and in combination with simulation techniques pointed to a dynamical mode crucial for the allosteric effect [75]. Neutron scattering has also been employed to elucidate structural origins of allostery, e.g. Li *et al.* demonstrated that ezrin binding to a scaffolding protein NHERF1 induces a conformation change that reaches across 120 Å and allostericaly regulates binding of C-CFTR to its two distant PDZ domains.

1.4.5 Fluorescence Resonance Energy Transfer (FRET)

Many fluorescence techniques have been developed and are suited for dynamics measurements and allostery analysis, most notable the fluorescence (Förster) resonance energy transfer (FRET) [76]. FRET is a process that occurs between two fluorophores, a donor and an acceptor. An energy source is used to excite the donor, which then transfers its energy via dipole-dipole coupling to the acceptor if they are sufficiently close to each other. The FRET signal is sensitive to distance between the donor and the acceptor and can provide accurate information on distances in the 1-10 nm range [77]. The limitation

of traditional FRET stems from background fluorescence of the assay components such as the buffer and the protein.

The background fluorescence is typically very short-lived (ns range) and can be overcome by time resolved FRET (tr-FRET). Additionally tr-FRET offers a mean to measure the protein flexibility. In a tr-FRET experiment the donor emission is measured after a short exciting pulse. In the absence of an acceptor this emission is very short (ps-ns range) and decays exponentially with the intrinsic life-time of the donor. In the proximity of the acceptor the decay becomes modified and the distance can be inferred. The measurements can be repeated with ns-μs frequency depending on the intrinsic lifetime of the donor's fluorescence. The data yield a distribution with the width reflecting the flexibility of the macromolecule. For details of the FRET techniques see e.g. [76, 78].

Steady state FRET is frequently used for folding studies [76] but has been also a key tool in some flexibility measurements, e.g. of nucleic acids [78, 79]. Allosteric proteins have been subjected to FRET studies too, e.g. Polit *et al.* investigated the two level allosteric mechanism of CAP [80] (see Chap. 3 below).

Chapter 2

Coarse-Graining Methods in Biophysics

In biophysics we use physical methods to shed light on problems in biology. The scientific method of physics consists of interaction between theory and experiments. A problem is understood from a physics point of view if a theory exists and its predictions have been verified experimentally. In the previous chapter we described many experimental techniques able to investigate allostery and probe protein dynamics. In this chapter we describe the theoretical methods suitable to confront this body of data.

Allostery and protein dynamics are classical problems of molecular biology. The ideal theoretical approach needs to take into consideration the atomistic details of the molecule. Computational methods such as molecular dynamics or Monte Carlo are in principle able to investigate molecules and their dynamics in atomistic detail but suffer from a common problem that they cannot simulate the system for sufficiently long periods of time. In practice this means that in a fully atomistic simulation the captured component of protein's dynamics is limited to the high frequency modes. The low frequency motions, crucial for allostery, are currently out of reach of atomistic simulations.

Centuries of research in natural phenomena suggest that some fine details of the atomistic structure may not be required to answer questions concerning low frequency motions. After all, these motions involve large, structurally compact structures that move in a concerted fashion. The role of a physicist is to identify the crucial features of the structure and construct a model at the appropriate degree of coarse-graining with the

redundant details averaged out. Such model is considered appropriate if it captures the phenomenon in question and also has predictive power.

In this chapter we discuss different levels of coarse-graining and what biophysics problems they are appropriate for. The finest and most accurate simulations use quantum theories to model interactions within the molecules. Since such modeling is limited to small molecules and very short time scales we do not consider it in this thesis. The next level of computational techniques has atomistic resolution. The most common method is molecular dynamics (MD), reviewed in greater detail below. A Monte Carlo technique is theoretically capable of simulating biomolecules in atomistic details but due to technical difficulties is rarely used in such studies [81]. It is more commonly employed in combination with more coarse-grained models of molecules, such as worm like chain model for DNA [82], protein folding [83] or bioinformatics [84]. Since we do not consider any results of Monte Carlo simulations in this thesis we do not detail this technique any further.

An MD simulation produces a trajectory of the molecule that contains a vast amount of information on the molecular dynamics. However, extracting a quantitative description of the dynamics is not trivial. Many methods have been developed for this purpose and we introduce some of them in Sec. 2.2.

There are two ways to overcome the problem with the short timescales accessed by the traditional molecular dynamics. Firstly, the biomolecule can be coarse-grained structurally. In this approach a certain number of atoms is grouped together and represented as a single point with extrapolated properties, such as mass, charge etc. In the other approach the fairly complex interaction potential of a typical MD simulation (see Eq.(2.1)) is simplified. In practice both approaches are used, often in combination. We discuss these more coarse-grained models in section 2.3 and 2.4 and focus on models used in this thesis. Finally we introduce the methodology developed by Hawkins and McLeish [38, 42] and used in this thesis, and explain where it fits in the broad picture.

2.1 Molecular Dynamics

Molecular dynamics (MD) is currently the most popular method for atomistic modeling of biomolecules. The result of an MD simulation is a trajectory of every atom of the

system during a set amount of time. The trajectories are found by solving Newton's equations of motion at discrete time steps.

The atoms of the system interact through an empirical pairwise potential commonly referred to as the force field. There are many molecular dynamics software packages available that differ in the implementation of the algorithm but the functional form of the force field is well preserved amongst all of them. The parameters for the field are found from quantum mechanical calculations and differ slightly between different programs. The most common form of the potential is composed of four terms [85]

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]$$
$$+ \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right]. \tag{2.1}$$

The first term sums over all covalently bonded atoms representing their bonded interactions by harmonic springs. The second term defines the bond angle potential that is dependent on the geometry of electron orbitals of the involved atoms. The third term sums over all torsional angles to account for energy of twisting of a bond. Finally the fourth term consists of Van der Waals and electrostatic potentials to account for the non-bonded interactions between all atom pairs. The most commonly used programs include CHARMM [86], AMBER [85], and GROMACS [87]; the references state the values of the parameters and describe the way they have been obtained.

Molecular dynamics requires an input of a three dimensional structure of the biomolecule. The structure is typically obtained with an experimental technique, such as X-ray crystallography or NMR. In the simulation the molecule is submerged in a solvent, which can be either implicit (the molecule is surrounded by a continuum with average properties of a solvent) or explicit where each solvent molecule is defined atomistically. At the start, every atom of the system is assigned a pseudo-random velocity so that the kinetic energy of the system corresponds to the set temperature through the equipartition theorem.

The velocity and the force obtained from integrating the force field is used to calculate the position of every atom at the next time step. This process is repeated at every time step resulting in large requirements on the computer power. The current MD simulations typically cover tens to hundreds of nanoseconds with a common time step of 1 fs. This is usually not long enough to capture the slowest modes.

Apart from the limitations connected with simulation time and system size the technique has another weakness: the empirical pairwise interaction potential. Within the potential the covalent bonds are represented by harmonic springs, a good approximation around the equilibrium position of the bond but very inaccurate when bonds are highly strained or breaking [88]. The second major approximation is the pairwise nature of the potential. The simplification of the many body interaction to pairwise usually results in a correctly reproduced structure but can lead to a wrong internal energy and thermodynamic values [89].

Despite of its shortcomings, MD is currently employed as a main investigation tool or as a complementary techniques in a large number of problems in molecular biology. The modern MD programs are equipped with a relative friendly user interface and the visualisation packages provide a window into the atomistic world of molecules rendering the technique exceptionally attractive. MD serves a number of tasks ranging from its microscope-like function to its role in molecular structure refinement. For our purposes the most important feature is the large amount of information about the dynamics of the studied molecule contained in the MD trajectory.

2.2   Analysis of a Simulated Trajectory

An MD trajectory can be visualised with molecular graphics software, e.g. VMD or Chimera used in this work. The software produces an animation of the molecular motion and flexible regions can be observed. However in order to extract quantitative description analytical tools have to be used. Our primary interests are thermodynamic properties (free energy, enthalpy and entropy), decomposition of the complex motions into uncorrelated components and mechanical properties, such as local rigidities. Here we review the major methods suitable for the thermodynamic and dynamic mode analysis. The protein rigidities are best extracted from the normal mode and principal component analysis. For DNA a different technique, known as the base pair step analysis is commonly used. Due to the extensive, DNA-specific introduction required and little general interest we explain this technique in chapter 4.

Allosteric reactions are thermodynamic in nature and thus the minimal quantitative description consists of the free energy differences of each binding step and its decomposition into enthalpic and entropic terms. However, calculation of the free energy and its

components from a computer simulation is an extremely difficult task. Despite substantial progress in this field most methods are limited with inaccuracies and inefficiency [90].

The main difficulty lies in the calculation of the absolute values of the thermodynamic parameters. In particular the evaluation of absolute entropies is a notoriously difficult. The situation is much improved when only the difference in free energy (entropy) between two microstates ΔF (ΔS) is sought after. However, the methods used in this case (e.g. the counting approach and the coupling parameter approach) are reliable only in cases when the two microstates lie very close on the free energy landscape and thus are not suitable for large structural and dynamical changes. Since this is the case for the majority of allosteric systems we concentrate mainly on the methods designed to determine the absolute values of F and S and only briefly review the methods for ΔF calculations. Other methods for ΔF determination, such as adaptive integration method or methods based on Jarzynski identity, are reviewed in [90].

2.2.1 Calculation of Relative Thermodynamic Parameters

2.2.1.1 The Counting Approach

The counting approach provides the most straight-forward method for the calculation of the free energy difference between two microstates. The amount of time a system spends in a certain microstate relative to the total simulation time is proportional to the probability of the microstate. The free energy difference between two microstates (1 and 2) can therefore be found by counting the number of visits in the particular microstates

$$\Delta F = k_B T \ln \frac{n_1}{n_2}, \tag{2.2}$$

where n_1 and n_2 is the number of times the simulation "visits" the state 1 and 2 respectively.

The obvious requirement for this method is that the MD trajectory is long enough to "visit" the microstates in statistically significant amounts. Ligand binding however occurs on very slow timescales rendering this method unsuitable for our purposes.

2.2.1.2 Thermal Integration (TI)

This method has similarly been designed for calculating differences in free energy and entropy between two states. The states have to lie in close proximity on the free energy landscape, but are not required to fluctuate naturally between the two as in the counting approach. In TI a parameter that couples the two states is found. The system is then driven along this reaction coordinate from one state to the other. The integration can be performed over a wide variety of parameters such as energy, temperature, pressure or non-physical quantities through so called "alchemical mutation" that allows small groups of atoms to be substituted [90]. The method is therefore suited for slightly different problems, e.g. calculation of the free energy difference between binding two ligands very similar to each other to an identical binding site [91] or the free energy penalty for introducing a charge [92].

Thermal integration has many variations which are summarised e.g. in the recent review of Meirovitch [90]. All the variations share the same basic principle, which is embodied in Zwanzig's formula

$$\Delta F = -k_B T \ln \left\langle \exp \left(\frac{-\Delta E(x)}{k_B T} \right) \right\rangle_1 , \qquad (2.3)$$

where ΔE is the energy difference between the two states and the index 1 signifies that the average is taken with respect to the first state. Energy is easy to measure in an MD simulation and therefore the equation is a very straight-forward way to obtain ΔF. The Zwanzig formula is only applicable though if ΔE is small, due to computational limitations. The usual trick is to divide the energy into little "windows" defined by the interval $(\lambda - \Delta\lambda, \lambda + \Delta\lambda)$ around values of reaction coordinate λ, running a simulation in each λ-window. The total ΔF is then obtained from Kirkwood's relation

$$\Delta F = \int\limits_{\lambda=0}^{\lambda=1} \frac{\partial \langle E(x,\lambda) \rangle}{\partial \lambda} \, \mathrm{d}\lambda \qquad (2.4)$$

The main advantage of this method is that it enables us to calculate small free energy differences in large systems.

Entropy can then be obtained in several ways, the most reliable is to run the simulations at several temperatures and use the relation

$$S = -\left(\frac{\partial F}{\partial T} \right)_V \qquad (2.5)$$

There are two obvious disadvantages of this method. First of all it is computationally very expensive as we need to run a whole MD simulation for each λ window. And secondly this method is practically feasible only in cases with small variance in structure and dynamics of the two systems.

2.2.2 Calculating Absolute Entropies

The approaches that calculate free energy and entropy difference are all hampered by the same problems: the two states have to be very close on the free energy landscape. This problem can in principle be overcome by calculating the absolute values of the thermodynamic parameters in each state. Whilst the enthalpy can be evaluated from the potential energy, the extraction of an (within the limits of the simulation methodology employed) exact value of the entropic component of the free energy is principally an impossible task, because it requires sampling of the whole phase space and thus an infinitely long simulation. Statistical thermodynamics comes to the rescue and offers a means to estimate the entropy from an incomplete sampling.

2.2.2.1 Quasiharmonic Analysis

The general approach is to estimate the entropy based on an analytical approximation of the conformational probability distribution of the system. The currently most employed method, the quasiharmonic analysis, known also as the principal component analysis, was proposed by Karplus and Kushick in 1981 [93]. The method assumes that the local potential function can be expressed as a quadratic function of the internal coordinates of the atoms. This leads to a multivariate Gaussian distribution of the configurational probability, a distribution associated with a classical harmonic oscillator.

The entropy of a one-dimensional harmonic oscillator in the classical limit ($k_B T \gg \hbar\omega$) is

$$S = k_B \left(1 - \ln \frac{\hbar\omega}{k_B T} \right), \tag{2.6}$$

where ω is the angular frequency of the oscillator.

Although MD is a classical approach, some of the MD frequencies, in particular the bond stretching and bond-angle bending vibrations, are high enough for the classical approximation to break down. The formula for entropy of a quantum mechanical harmonic

oscillator [94], or more commonly a semi-classical approximation derived by Schlitter [95] therefore has to be used.

The Schlitter semi-classical approximation of entropy reads

$$S = \frac{k_B}{2} \ln \left(\det \left[\mathbf{I} + \frac{k_B T e^2}{\hbar^2} \boldsymbol{\sigma}' \right] \right), \tag{2.7}$$

where $\mathbf{I}$ is the identity matrix and the mass weighted covariance matrix $\boldsymbol{\sigma}'$ is defined as

$$\boldsymbol{\sigma}' = \mathbf{M}^{1/2} \boldsymbol{\sigma} \mathbf{M}^{1/2}. \tag{2.8}$$

The inertia matrix $\mathbf{M}$ is a diagonal matrix with the elements $M_{1,1} = M_{2,2} = M_{3,3} = m_1, \ldots, M_{3N,3N} = m_N$, where m_i denotes the mass of the atom i.

The covariance matrix $\boldsymbol{\sigma}$ is defined as

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{2.9}$$

where $x_1, \ldots, x_{3N}$ are the Cartesian coordinates of the atoms.

In the article Schlitter also demonstrated that this formula provides an upper bound of the real entropy of the system.

The entropies obtained by the Schlitter method have been shown to produce inconsistent results in some cases, predominantly due to the incorrectly treated anharmonicities and correlations among the probability distributions. For a detailed discussion on this topic and a suggested partial solution see [96].

The mass weighted covariance matrix Eq. (2.8) can also be used for another purpose: principal component analysis (PCA). Diagonalisation of the matrix yields eigenvalues λ_i that are related to the quasiharmonic frequencies $\omega_i = \sqrt{k_B T / \lambda_i}$. Six of these frequencies correspond to the translation and rotation of the whole molecule and are therefore zero. The remaining eigenvalues can be used to find the eigenvectors and to reconstruct the quasiharmonic modes around the average system conformation.

Harris and Laughton have shown that PCA provides a very good reduced representation of the dynamics of DNA [97]. The highest frequency modes correspond to the bond vibrations and the slowest modes are reminiscent of normal modes, such as bending and twisting of the entire helix. DNA is known to behave as a semiflexible polymer [98] and as

such has a relatively smooth free energy landscape. The application of PCA to globular proteins is much more complicated. Proteins contain highly flexible as well as very rigid regions and the slow and fast modes are nontrivially coupled [99]. Together with the complication of short simulation timescales this may result in a poor representation of the dynamics [100]. Quasiharmonic analysis has been used by Harris *et al.* to study dynamic allostery in DNA [101]. They concluded that the dynamics is entropy driven and that all modes contribute to the total effect. PCA was used to illustrate how a drug binding affects the global modes. More details on their results is given in Chap. 4.

Stacklies *et al.* used principal component and quasiharmonic analysis to investigate the mechanism of dynamic allostery in methionine receptor [102]. The authors identified low frequency cooperative modes and quantified their changes upon binding. They also used the Schlitter formula to calculate entropy but with mixed results. The values of entropy are qualitatively correct: binding of a ligand always decreases the entropy of the repressor, but the presence of DNA leads to a 20 fold smaller value. This is considered a rather large and probably unphysical result [102].

2.2.2.2 Normal Mode Analysis

MD in combination with quasi-harmonic analysis are computationally expensive and are thus unsuitable for many larger systems. One way to characterise large systems and/or to speed up the procedure of dynamic analysis of the smaller system is to use normal mode analysis (NMA). Normal mode analysis assumes that the displacement of the atoms around their equilibrium position is small enough that the interaction potential (Eq. (2.1)) can be approximated by a harmonic potential. In more formal language this corresponds to performing the Taylor expansion of the potential around the equilibrium position and then neglecting third and higher order terms. The complex motion of the molecule can then be decomposed into a set of decoupled normal modes.

A standard NMA is a two step process. In the first step, MD is used to find a local minimum on the free energy landscape. This step may be time consuming but constitutes a substantial reduction of computer time compared to the PCA technique. In the second step, a quadratic expansion around the minimum is found. This allows the Newton's equation of motion to be written as

$$\mathbf{M}\ddot{\mathbf{x}} = -\mathbf{H}\mathbf{x}, \tag{2.10}$$

where $\mathbf{x}$ are the Cartesian coordinates of the atoms in the local minimum, $\mathbf{M}$ the mass matrix defined in 2.2.2.1 and $\mathbf{H}$ the Hessian matrix defined as

$$H_{ij} = \frac{\partial^2 E_{total}}{\partial x_i \partial x_j}, \tag{2.11}$$

evaluated at the local minimum. The potential E_{total} is the typical potential function of molecular dynamics (Eq. (2.1)) but can in principle be any intraction potential.

The mass weighted Hessian matrix $\mathbf{H'} = \mathbf{M^{-1/2}HM^{-1/2}}$ is diagonalised to yield eigenvalues related to the frequencies of the normal modes. The first six eigenvalues equal zero and correspond to the translation and rotation of the whole molecule. The eigenvectors of $\mathbf{H'}$ define the motion of the atoms within the individual normal modes.

The reconstructed normal modes can be used to yield a wide range of interesting parameters such as the crystallographic B-factors, RMS fluctuations or dynamic correlation maps [103].

The frequencies obtained with NMA can also be used to obtain the entropy of the system. However, NMA only probes vibrations associated with a single structure, i.e. a very small region of the phase space. A good estimate of the entropy can only be obtained by sampling a substantial part of accessible configurational space. NMA can only access a significant region of the phase space of a very stiff and thus highly harmonic molecule. Solvated macromolecules however do not belong to this category and the quasi-harmonic

Quasiharmonic
approximation

Normal Mode
Analysis
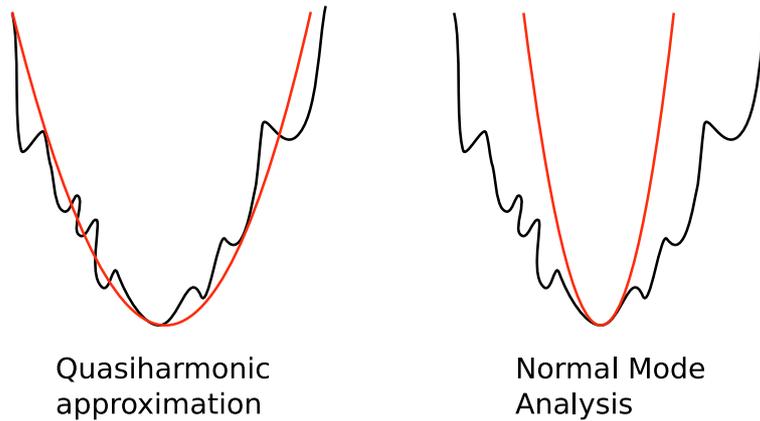
FIGURE 2.1: A schematic comparison of quasiharmonic approximation and normal mode analysis. The rugged free energy landscape (black) is better approximated by the quasiharmonic potential.

analysis is a more suitable tool for the entropy determination. For schematic comparison of the two methods see Fig. 2.1.

2.3   Elastic Network Models

Although not suitable for entropy calculations in biological problems, normal mode analysis (NMA) provides an illustrative and very useful description of the normal modes of the molecule. Collaborative efforts of the biological community resulted in a very useful database of known structures of different biomolecules. It has become recognized that it would be desirable to complement this database with information on the molecules' dynamics. Standard NMA however requires too much computer time to make it feasible to obtain such description of all known macromolecules.

In 1996 Tirion proposed a revolutionary solution to this key problem of the field. She noticed that the complex MD potential could be substituted by a single-parameter harmonic potential and most importantly that when this is done, the low frequency modes remain nearly unaffected [104]. Furthermore she observed that these collective motions are insensitive to detailed molecular structure and therefore the minimization previously required before the normal mode analysis can be avoided. This constitutes a radical reduction in computer time and enables full dynamic analysis to be performed in order of minutes.

The Tirion's pairwise potential is defined as

$$E(\mathbf{r}_i, \mathbf{r}_j) = \frac{C}{2} \left(|\mathbf{r}_{i,j}| - |\mathbf{r}_{i,j}^0| \right)^2 . \tag{2.12}$$

The vector $\mathbf{r}_{i,j} = \mathbf{r}_i - \mathbf{r}_j$ denotes a vector connecting atoms i and j, the superscript 0 the initial configuration. The constant C is an empirical constant and is identical for all atom pairs rendering the evaluation of the Hessian matrix substantially faster.

The total potential energy is found as a sum of the energies from Eq. (2.12) over all pairs of atoms separated by less than a cutoff distance. Tirion argues that the best results are obtained for low cutoff values and recommends the use of 2 Å. The value of the phenomenological constant C is adjusted so that $CR^2 \approx k_B T$, where R is the cutoff distance.

Alternative potentials have been proposed, e.g. an exponential potential proposed by Hinsen [105], and have been shown to yield very similar results. In fact, several studies showed that slow, collective motions of a protein are insensitive to detailed molecular structure and the force field so long the large-scale geometry of the structure is accurately modeled. The higher frequency modes are more sensitive to the details of the structure and potential and are thus not well represented in the elastic network model.

Since its original discovery, several groups worldwide took advantage of the speed and simplicity of this method and released software that performs the NMA analysis with different coarse-grained potentials. Many of these programs are available on web servers open to the public so that anybody can submit a protein structure and within a few minutes obtains a full analysis. There are only very few servers that perform the analysis as originally suggested by Tirion, e.g. NOMAD-Ref [106]. The technique has developed since with new implementations of the Hessian matrix and more coarse-grained versions. We mention these alongside the explanation of the techniques over the next paragraphs. An extensive review of large number of the public web instruments has been recently published by Liu and Karimi [107].

2.3.1 Gaussian Network Model

Bahar *et al.* took the coarse-graining a step further by showing that the lowest frequency modes are well represented in an elastic network model where the protein is modeled by a reduced set of mass points each representing a whole residue [108]. The mass points are located in place of C_α atoms (backbone atom attached to each residue) and are connected with identical springs. For the implementation of the spring network the authors took inspiration in Flory's theory of polymer networks, where junctions of the network undergo Gaussian-distributed fluctuations [109]. This version of the elastic network model became known as the Gaussian network model (GNM).

In the protein model the Flory junctions (here the C_α atoms) are connected by springs with a uniform force constant γ if closer than a cutoff distance. An example of such representation is shown in Fig. 2.2. The exact form of the potential is

$$U = \frac{\gamma}{2} \sum_{i,j} (\mathbf{r}_{i,j} - \mathbf{r}_{i,j}^0) \cdot (\mathbf{r}_{i,j} - \mathbf{r}_{i,j}^0), \tag{2.13}$$
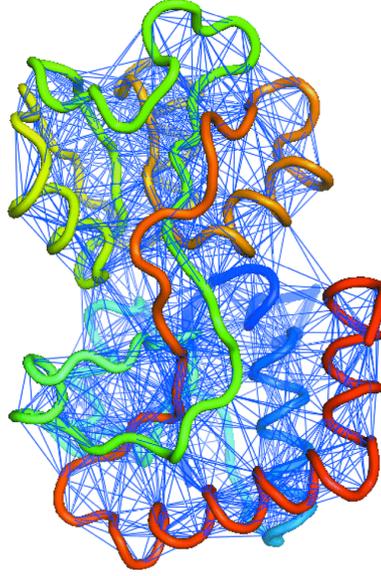
FIGURE 2.2: Glutamine-binding protein in the elastic network model representation. Adapted from [110].

where $\mathbf{r}_{i,j} = \mathbf{r}_i - \mathbf{r}_j$ denotes a vector connecting atoms i and j, the superscript 0 refers to the initial configuration.

The principle of constructing the network is identical to the ENM but the Hessian matrix is replaced by the Kirchhoff or connectivity matrix $\mathbf{\Gamma}$. The elements of the Kirchhof matrix are found as follows

$$\mathbf{\Gamma}_{ij} = \begin{cases} -1 & \text{if } i \neq j \text{ and } r_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } r_{ij} > r_c \\ -\sum_{k, i \neq k} \mathbf{\Gamma}_{ik}, & \text{if } i = j \end{cases} \tag{2.14}$$

where r_{ij} is the distance between the atoms i and j, and r_c is the cutoff distance. The neighboring C_α atoms are $\sim 3.8\,\text{Å}$ apart and the cutoff distance is typically chosen as $7\,\text{Å}$.

The inverse of the Kirchhoff matrix is related to the equilibrium correlation between fluctuations $\Delta\mathbf{r}_i$ and $\Delta\mathbf{r}_j$

$$\langle \Delta\mathbf{r}_i \cdot \Delta\mathbf{r}_j \rangle = \frac{k_B T}{\gamma} [\mathbf{\Gamma}^{-1}]_{ij} \tag{2.15}$$

This allows for a direct comparison with the B-factor with γ as the only fitting parameter. The B-factors obtained with this method were found to agree well with the values reported from the X-ray crystallography [111]. GNM calculates only amplitudes and frequencies of the normal modes but does not provide information on the direction of motion.

GNM is implemented on the web server iGNM (http://ignm.ccbb.pitt.edu). iGNM calculates 20 slowest, 20 fastest modes, correlation between fluctuations and performs several other analyses (for details see [112]).

NMA has recently been implemented with single site per residue level of coarse-graining too. Web servers that provide such coarse-graining include WEBnm (http://services.cbu.uib.no/tools/normalmodes) and AD-ENM/DC-ENM (http://enm.lobos.nih.gov).

It soon became clear that the coarse-graining can be taken even further for large systems. NMA computation time scales as N^3 where N is the number of beads representing the structure. Thus a reduction of the number by an order of magnitude reduces the computation time by three orders of magnitude. Several coarse-graining methods have been developed [113]; here we review only the most popular rotation-translation-block (RTB) method implemented at the web server ElNemo.

2.3.2   Rotation Translation Block Approximation

Tama *et al.* performed a systematic analysis of the low frequency modes obtained by different levels of coarse-graining [114]. Based on a set of 12 proteins with sizes ranging between 46 and 858 residues, they concluded that when 6 amino acids are coarse-grained into a single rigid block, the normal modes are nearly identical to the modes obtained from all atom NMA.

In the rotation translation block (RTB) approach the Hessian matrix required for the normal mode calculation is expressed in a new set of coordinates, the rotations and translations of the n_b blocks. The projected Hessian is given by

$$\mathbf{H_b} = \mathbf{P}^t\mathbf{H}\mathbf{P} \tag{2.16}$$

where $\mathbf{P}$ is a $3N \times 6n_b$ matrix containing the rotations and translations of each block and $\mathbf{H}$ is the all atom $3N \times 3N$ Hessian matrix. Diagonalization of $\mathbf{H_b}$ yields the frequencies of the normal modes and the atomic displacements are given by

$$\mathbf{A_p} = \mathbf{PA_b}, \tag{2.17}$$

where $\mathbf{A_b}$ is the matrix of eigenvectors of $\mathbf{H_b}$.

The RTB method is implemented on the server ElNemo (http://igs-server.cnrs-mrs.fr/elnemo/) which calculates up to 100 slowest modes. It outputs the eigenvalues and eigenvectors together with a graphical representation of the modes. The website offers several options for more detailed analysis of the results, details can be found in [115].

The simplified potential of elastic network models provides a very attractive opportunity to study protein dynamics. It enables a full dynamic analysis to be performed in order of seconds to minutes and in combination with coarse-graining techniques such as the RTB poses practically no limitations on the system size. Moreover, it has been confirmed many times since the original proposition that the low frequency modes are in excellent agreement with other techniques and experiments [104, 105, 113].

The low frequency modes are nearly independent of the level of coarse-graining and the interaction potential suggesting that the contact topology determines the nature and the frequency of these functional modes and not the detailed chemistry. Assuming that proteins have evolved to an optimal biological function leads to the conclusion that these robust modes must carry a part of the functional tasks. The relative ease of describing these modes only underlines the need to understand what role they play in the cell and through what intricate mechanisms.

ENM has been used on numerous examples in the allostery investigations. Williams uses ENM to study protein kinase PDK1 and finds that effector binding activates correlated motions in the region of the substrate binding site [116]. Keskin *et al.* use the GNM to investigate the allosteric mechanism in the GroEL-GroES complex. They identify the motions that are most affected by the effector binding and thus shed light onto this extremely long range ($> 100\,\text{Å}$) allosteric mechanism [11].

Elastic network models are predominantly used for protein dynamics analysis but can also be used for DNA. Only two servers offer a model of nucleic acids, in particular iGNM and AD-ENM/DC-ENM. On these servers each nucleotide is represented by 1-3 beads that are connected by uniform springs. These coarse-grained models are however typically only applied to DNA-protein complexes mainly because the motion of DNA on a small scale is better described by quasi-harmonic approximation and on a large scale the worm like chain model. For details see chapter 4.

Several attempts have been made at improving the elastic network models, most notably Ming and Wall suggest that strengthening interactions between backbone neighbors significantly improves the density of states distribution, in particular the high frequency C_α vibrations [117]. Since the elastic network models are mainly employed for the low frequency mode determination this model has never become widely used in the community.

2.4   FIRST and FRODA

The elastic network models we have introduced so far use coarse-grained representations of proteins consisting of arbitrarily selected atoms along the backbone (e.g. every C_α atom) justifying the choice by showing that the structure of the lowest modes is independent of the selection (e.g. [113, 114]). FIRST/FRODA adopts a different approach where atoms that naturally form rigid regions are grouped together to form the basis of the coarse-grained mechanical model. A Monte Carlo geometrical simulation is then performed with this model to generate a trajectory in a conformational space of the molecule.

FIRST (Floppy Inclusions and Rigid Substructure Topography) is an algorithm that offers a rigorous way of finding rigid domains of a protein. The technique, first introduced by Jacobs *et al.* in 2001, uses tools from graph theory to analyze a bonded network obtained from the 3D structure of the protein [118]. The protein is modeled as a network where atoms correspond to the vertices of the graph and edges represent bonds. Only covalent, hydrogen and salt bridges bonds are considered. The bond distances and angles are kept fixed (constrained) and only dihedral angles constitute degrees of freedom. This type of network is known in mathematics as the bond-bending network and its stability can be determined [118].

FIRST uses the so called pebble game to determine which bonds are rotatable within the defined constraints. The pebble game is an algorithm that simplifies the search from $O(N^2)$ to linear scaling; rendering it fast enough for every day use (an average protein can be analysed in the order of seconds). Rigid clusters are found by counting the number of constraints and degrees of freedom of all bonds. When the number of constraints exactly matches the number of degrees of freedom such region is called isostatically rigid, if there are abundant constraints the region is over-constrained or stressed, otherwise the region is flexible. The result of the FIRST analysis is the division of the protein into rigid (stressed) and flexible regions the size of which depends on the user-defined energy cutoff that determines which bonds to include. Rigid regions contain anywhere between three and hundreds of atoms. The cutoff ranges from 0 to -10 kcal/mol, the default value is -1 kcal/mol.

FIRST determines the flexible and rigid regions statically; it does not determine the motion but the potential to move. FRODA (Framework Rigidity Optimized Dynamic Algorithm) has been developed to examine the actual motions of the coarse-grained protein. It performs a geometrical simulation where conformers are generated within the defined constraints using Monte Carlo. Trajectories in conformational space are obtained instead of dynamical trajectories [119].

FIRST and FRODA are freely available on the web server Flexweb (http://flexweb.asu.edu).

2.5 Our Methodology

The coarse-graining method we use in this work has been developed by Hawkins and McLeish [38, 41, 42]. It is motivated solely by the objective to explain the role of thermally excited motions in allosteric effects.

Thermal fluctuations of the macromolecular native environment excite a whole spectrum of internal vibrations. Of these, mainly the slow internal motions are believed to contribute to long-range allosteric signaling [19]. A slow, global mode involves a whole structural unit such as a helix or a domain. The perturbation of such motion therefore directly influences distant binding sites. Conversely, fast motions are localized and consequently only affect a few atoms within their localization length.

The slow global modes are practically impossible to observe with all atom simulations. Yet from the study of elastic network models we learned that the slowest modes are very robust, i.e. independent of the atomistic detail and their interaction potential and can thus be simulated with coarse-grained techniques. In particular we know that harmonic potentials and rigid blocks containing several residues provide a good enough approximation and result in slow modes nearly identical to those obtained from fully atomistic molecular dynamics [114].

Although ENMs may seem an ideal tool for studying allostery they have a few drawbacks. Firstly they require crystal structure for all binding states and these are often not available. Furthermore we show in chapter 3 that on occasions where there is very little structural change activated by a small ligand binding to the macromolecule, elastic network models fail to observe experimentally reported substantial changes in the macromolecular dynamics.

Because of the limitations of this class of theoretical models Hawkins and McLeish chose a different methodology and constructed models aimed at explaining the allosteric behavior in systems with limited structural change [38, 41, 42]. The objective was to find an appropriate level of coarse-graining, ideally simple enough to allow for analytical treatment, and yet able to capture the desired allosteric effect. They selected a class of proteins with relatively obvious low frequency modes and then systematically found the appropriate model. This model is for obvious reasons much coarser than any of those introduced above.

In the model whole domains are treated as rigid bodies that interact through a harmonic potential. As an example a model of the lac repressor is shown in Fig. 2.3. Lac repressor is composed of two relatively rigid identical monomers connected with soft hinges. In the model of Hawkins and McLeish the repressor is coarse-grained into two rigid plates representing each monomer, connected by a minimal set of harmonic potentials forming the interaction between the monomers. The model of the repressor has only six degrees of freedom corresponding to the lowest six modes of the protein. Binding of a ligand is modeled as a change of stiffness of the local spring constant.

The authors have demonstrated that the six global modes can contribute significantly to the allosteric free energies and that the model describes both positive and negative cooperativity in a general dimer of rigid monomers. They successfully parameterised this
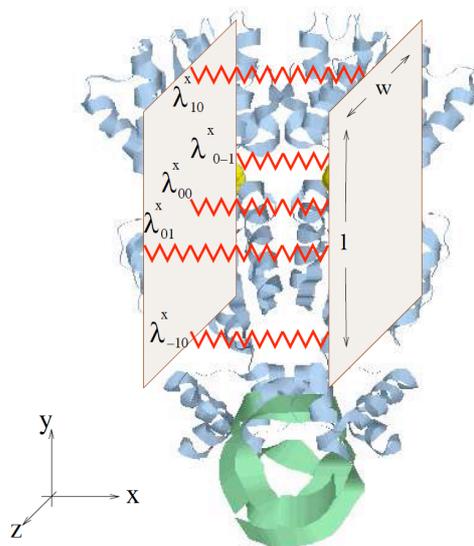
FIGURE 2.3: Lac repressor dimer coarse-grained into two rigid plates corresponding to the two monomers and a minimal set of springs representing the interaction between the monomers. The repressor is shown bound to DNA (in green). Adapted from [38].

model from simulations and experiments for the case of lac repressor and thus presented a protocol for parameterisation of the model for other dimers [38].

The two coarse-grained models introduced in this thesis are also built for general classes of macromolecules, specifically symmetric homodimers and DNA oligomers. For the homodimers we follow the somewhat ad hoc coarse-graining method introduced by Hawkins and McLeish and confirm the appropriateness of the model retrospectively. In the case of DNA we first adopt a generally accepted coarse-grained model: an elastic rod. However, we find this model insufficient to explain positive cooperativity observed in DNA and have to extend the model in order to account for it.

Regardless of the exact form of the coarse-grained model the thermodynamic quantities are calculated in the following way. The system is described by a Hamiltonian H with a harmonic potential related to the chosen springs

$$\mathcal{H} = \frac{1}{2}\mathbf{p^T}\mathbb{M}^{-1}\mathbf{p} + \frac{1}{2}\mathbf{x^T}\mathbb{K}\mathbf{x} \tag{2.18}$$

where $\mathbf{p}$ and $\mathbf{x}$ are the momentum and coordinate vectors, $\mathbb{M}$ is the inertia matrix and $\mathbb{K}$ the elasticity matrix. We assume that the variations in mass are negligible during the binding and we can thus leave the inertia part of the Hamiltonian out of all our

calculations. The elements of the elasticity matrix are related to the spring constants and generally change upon the ligand binding.

In the next step we calculate the partition function using the theorem for multidimensional gaussian integrals. For a symmetric positive definite matrix $\mathbb{A}$ this theorem says

$$\int e^{-\frac{1}{2}\mathbf{x^T}\mathbb{A}^{-1}\mathbf{x}}d^n x = ((2\pi)^n \mid \mathbb{A} \mid)^{\frac{1}{2}}. \tag{2.19}$$

The partition function is then

$$Z = \int \ldots \int \mathrm{d}x_1 \ldots \mathrm{d}x_n \, \mathrm{d}p_1 \ldots \mathrm{d}p_n \, \exp\left(-\frac{H(x_1, \cdots, x_n, p_1, \cdots, p_n)}{k_B T}\right) = \frac{(2\pi k_B T)^n \mid \mathbb{M} \mid^{\frac{1}{2}}}{\mid \mathbb{K} \mid^{\frac{1}{2}}} \tag{2.20}$$

and the free energy

$$G = -k_B T \ln Z = \frac{1}{2} k_B T \ln |\mathbb{K}| + const., \tag{2.21}$$

where $|\mathbb{K}|$ denotes the determinant of $\mathbb{K}$.

We perform this calculation for each binding state and evaluate the allosteric free energy, a suitable measure of the cooperativity (Sec. 1.2). In this thesis we only investigate homotropic allostery, i.e. the ligands associating with the macromolecule are identical. In this case the allosteric free energy is defined as

$$\Delta\Delta G = (G_2 - G_1) - (G_1 - G_0), \tag{2.22}$$

where G_0 is the free energy of the free protein, G_1 of singly and G_2 of doubly bound. If the first binding inhibits the second, i.e. in case of negative cooperativity the resulting allosteric free energy is positive and vice versa.

Substituting Eq. (2.21) into (2.22) results in

$$\Delta\Delta G = \frac{1}{2} k_B T \ln \frac{|\mathbb{K}_0||\mathbb{K}_2|}{|\mathbb{K}_1|^2}. \tag{2.23}$$

The allosteric entropy can be obtained from standard relations between thermodynamic potentials

$$\Delta\Delta S = -\left(\frac{\partial(\Delta\Delta G)}{\partial T}\right)_p. \tag{2.24}$$

Consequently if the allosteric free energy is proportional to temperature the effect is purely entropic. Enthalpy is given by

$$H = k_B T^2 \frac{\partial \ln Z}{\partial T}. \tag{2.25}$$

The underlying principle of dynamic allostery can be understood from Eq. (2.21), which relates the free energy logarithmically to the effective stiffness. Let us explain the principle on a simple example of two identical ligands binding to the protein each of which decreases the flexibility of a single mode. The situation is sketched in Fig. 2.4. The ligand binding affects a single internal breathing mode of the protein pictured as a scissor like motion of two rods. We note that the scissor like motion has been chosen arbitrarily only for the illustration purpose. The general result will be independent of the actual geometry of the global mode. Each binding modifies the local constant k by a factor $\alpha > 1$. This stiffening is reflected in the free energy penalty, which as we will see, is always larger for the first step due to the logarithmic nature of the free energy.

The situation is pictured in Fig. 2.5 a. Tightening of a stiff protein is entropically cheaper than that of the soft protein, a crucial result for dynamic allostery.

More formally the allosteric free energy connected with this twofold loss of conformational entropy is calculated from Eq. (2.23) and yields

$$\Delta \Delta G = \frac{1}{2} k_B T \ln \frac{2k(2\alpha k)}{(k + \alpha k)^2} = \frac{1}{2} k_B T \ln \frac{4\alpha}{1 + 2\alpha + \alpha^2} \quad < 0 \tag{2.26}$$

The resulting free energy is negative for all values of α (Fig. 2.5 b), corresponding to positive cooperativity.
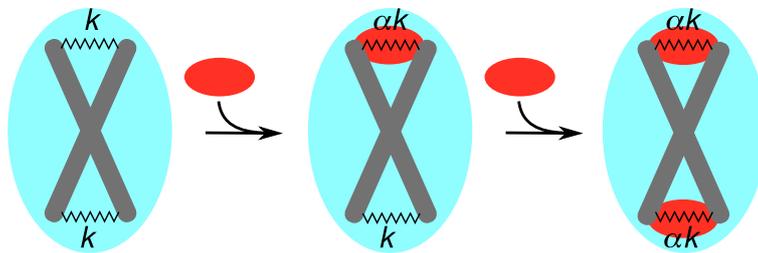
FIGURE 2.4: A sketch of a basic example of homotropic allostery. The ligand (in red) binding affects a single internal breathing mode of the protein (in blue) represented by scissor like motion of two rods. Each binding modifies the local spring constant by factor α.

(a)

(b)
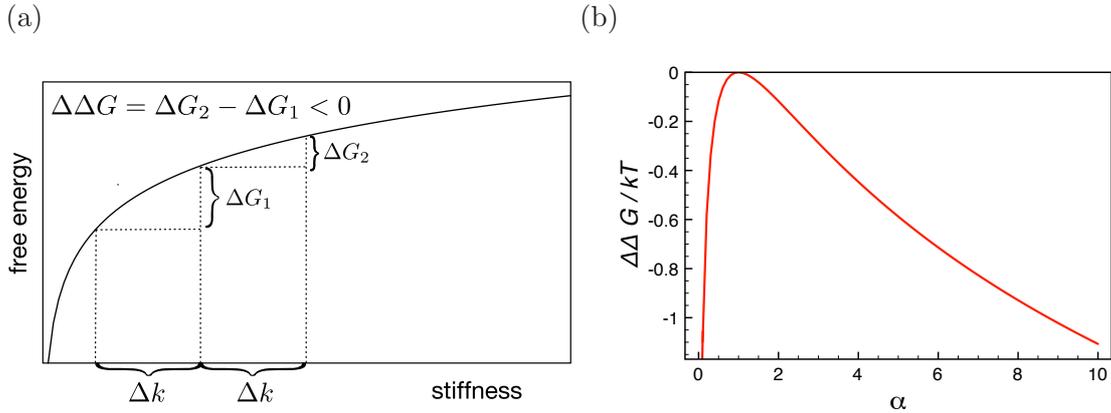
FIGURE 2.5: (a) Logarithmic dependence of free energy on the stiffness results in negative $\Delta\Delta G$ for the case of sequential identical tightening. (b) Allosteric free energy of the simple example of homotropic allostery shown in Fig. 2.4 as a function of enhancement parameter α.

A rise of negative cooperativity as a consequence of modified fluctuations is more subtle and will be discussed in greater detail in this thesis.

As mentioned previously, fast modes, such as side-chain movements, are typically localised in proteins [120], and consequently only affect a few residues within their localization length. However, as Hawkins and McLeish pointed out they can couple to slow modes and become involved in the communication indirectly [42].

The reasoning that fast modes may be involved in dynamic allostery was largely motivated by experimental results on the methionine repressor MetJ showing that despite the fact that MetJ displays allosteric behavior without a significant conformational change, large compensatory entropic and enthalpic allosteric energies are observed [121, 122]. In particular the enthalpic energies cannot be accounted for either by the associated structural changes or existing (slow mode) model of Hawkins and McLeish [42].

Hawkins and McLeish studied the consequences of coupling fast modes to the slow modes and concluded that the coupling can in a special case result in large compensatory entropic and enthalpic allosteric energies. Let us go through their calculation in more detail since we will use their approach on two occasions in this thesis.

We can picture the situation as shown in Fig. 2.6. The slow global mode of the protein is represented as a scissor-like movement of the two rods. Fast motions of smaller structures such as side chains are represented as vibrations of little protrusions attached to the rods.
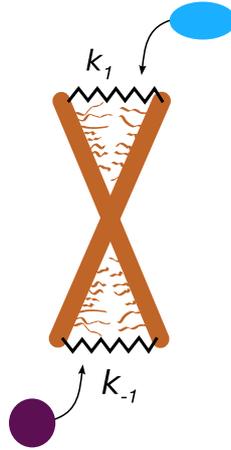
FIGURE 2.6: Schematic drawing of enslaved fast modes. Slow mode is pictured as a scissor like movement of the rods controlled by spring constants k_1 and k_{-1}. Little protrusions fluctuate at higher frequencies but can become coupled by the slow mode. The two rods can represent e.g. two alpha helices and the little protrusions side chains. Adapted from [42].

The coupling is based on the idea that the flexibility of the fast modes increases with the amplitude of the slow mode. Physically, local structures have more room for movement when their environment is disrupted, i.e. when the large rods are far apart. We assume therefore that the rigidity of the fast mode depends on the displacement $|x_s|$ of the larger structure within the slow mode. If $|x_s|$ is small, the localized structures are in their native environment, experience a deep, narrow potential, and move only slightly about the equilibrium position. If the slow mode becomes more flexible and thus $|x_s|$ larger, the fast mode environment becomes disrupted, and the corresponding potential becomes flatter.

Mathematically this idea is expressed by modifying the Hamiltonian of the system

$$\mathcal{H} = \mathcal{H}_s + \sum_{i=1}^{N} V_{f_i}(x_{f_i}, x_s), \tag{2.27}$$

where $\mathcal{H}_s$ is the Hamiltonian of the slow, global mode (scissor motion) and the sum adds up the contributions from N fast modes. The slow mode Hamiltonian is found from Fig. 2.6

$$\mathcal{H}_s = -V_{s_0} + \frac{1}{2}(k_1 + k_{-1})x_s^2, \tag{2.28}$$

where the spring constants k_1 and k_{-1} are affected by the effector and the substrate binding and V_{s_0} is the minimum of the slow potential.

The fast modes without any coupling are assumed to be harmonic too, so that

$$V_{f_i}(x_{f_i}) = -V_{f0} + \frac{1}{2}k_{f_i}x_{f_i}^2, \tag{2.29}$$

where V_{f0} is the minimum of the fast potential, here for simplicity assumed to be identical for all fast modes.

If the fast mode is coupled to the slow this potential becomes

$$V_{f_i}(x_{f_i}, x_s) = -V_{f0}f_i(x_s) + \frac{1}{2}g_i(x_s)k_{f_i}x_{f_i}^2, \tag{2.30}$$

where $f_i(x_s)$ and $g_i(x_s)$ are arbitrary real continuous functions that couple the fast modes to the slow. We choose these functions so that they mathematically define our physical requirements on the coupling, i.e. so that the fast mode potential becomes shallower and flatter with the increasing $|x_s|$.

Hawkins and McLeish made the following choice of the functions f and g, assuming them identical for all fast modes

$$f(x_s) = \frac{1}{1 + \frac{k_v x_s^2}{2k_B T}} \qquad\qquad g(x_s) = \frac{1}{1 + \frac{k_k x_s^2}{2k_B T}}. \tag{2.31}$$

The coupling parameters k_v and k_s have dimensions of force constants. In this thesis we prefer slightly different functions for the reason of analytical simplicity. The functions used in this thesis are following

$$f(x_s) = -\frac{k_v x_s^2}{k_B T} + 1 \qquad\qquad g(x_s) = \frac{1}{\exp\frac{k_k x_s^2}{2k_B T}}. \tag{2.32}$$

In order to calculate the partition function with this complex Hamiltonian Hawkins and McLeish used the method of steepest decent. In case of functions (2.32) no approximation is required to calculate the partition function. Hawkins and McLeish evaluated the allosteric free energy for two cases. Firstly, for $f(x_s) \equiv 1$, corresponding to constant fast potential depth, the authors showed that assuming $N/2 \gg k_s/k_k$ the allosteric free energy equals

$$\Delta\Delta G_{coupl} = (N+1)\Delta\Delta G_{slow}. \tag{2.33}$$

In other words the allosteric free energy arising from slow modes only is amplified $(N+1)$ fold by enslaving N fast modes.

If both the minumim of the fast potential and its width are affected by the slow mode, that is if $f(x_s) \neq 1$ and $g(x_s) \neq 1$, the resulting free energy is

$$G = \frac{1}{2} k_B T \ln \left(k_s + N \left(\frac{V_{f_0} k_v}{k_B T} - \frac{k_k}{4} \right) \right) + \text{const.} \tag{2.34}$$

Hawkins and McLeish assume $V_{f_0} k_v / k_B T \geq k_k/4$. If $V_{f_0} k_v / k_B T > k_k/4$ the resulting absolute allosteric free energy is smaller than in non-enslaved case. If $V_{f_0} k_v / k_B T = k_k/4$ the resulting allosteric free energy is identical to non-enslaved result but this energy contains enthalpic and entropic terms affected by the enslaved modes. The case of $V_{f_0} k_v / k_B T < k_k/4$ is studied in Chap. 5. Enthalpy is found from Eq. (2.25) and in the general case yields

$$H = \frac{1}{2} \frac{N k_v V_{f_0}}{k_s + N \left(\frac{V_{f_0} k_v}{k_B T} - \frac{k_k}{4} \right)} + \text{const.} \tag{2.35}$$

Clearly even for the special case of $V_{f_0} k_v / k_B T = k_k/4$ this enthalpic contribution is substantial.

Thus the enslaving of fast modes can result firstly in substantial amplification of the allosteric free energy but also in compensating enthalpic and entropic terms seen in the example of the Met repressor. We apply this type of coupling in both our models - the homodimer and the DNA model. Details of the coupling functions and the results are shown in the individual results chapters.

Throughout this thesis we use the classical harmonic approximation of the potential between the interacting coarse-grained structures. The classical approximation has been shown to break down for individual covalent bond vibrations, but these fluctuate at significantly higher frequencies than the structures we focus on. The choice of a harmonic approximation is on one hand the most straightforward and has been shown to work well for elastic network models but is it valid for the larger coarse-grained structures? Hawkins and McLeish tested it on the example of lac dimer with a technique based on the fully atomistic MD. Starting with the crystal structure of the dimer they fixed all the atoms in each monomer in order to emulate the rigid body approximation and then allowed the subunits to move with respect each other. The motion was performed in incremental steps and the energy for each structure was recorded. The result corresponds to the potential between the subunits and is very close to the exact harmonic approximation (Fig. 2.7). They studied nine lowest frequency modes and all displayed the same quadratic dependence but with different spring constants.
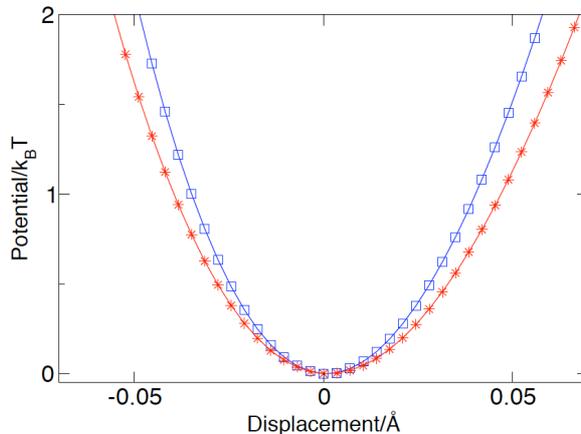
FIGURE 2.7: An example measurement of potential energy between two subunits of lac repressor. The two curves show the difference in spring constant of the repressor with and without bound inducer. Both curves are nearly parabolic corresponding to approximately harmonic potential between the subunits. Adapted from [123].

Another point worth noting is that the slow modes we study are not vibrational but over-damped. The overdamped motions are however repeatedly thermally excited. Thermo-dynamically we can treat the constantly re-excited modes in the same way as vibrational modes since their entropy is identical.

We outlined the coarse-graining procedure we apply in this thesis. The models are very general, meaning they can be in principle applied to any molecule from a particular target class (homodimers or DNA oligomers). The way to make these models specific is through their parameterization. This can be achieved through several methods. We mentioned that Hawkins and McLeish used crystallographic B factors, the technique based on all atom potential and elastic network models. Similarly we use the elastic network model and results from NMR for the homodimers and fully atomistic simulations for DNA. The details of the parameterization techniques are given in the corresponding results chapters.

Until now we have been discussing three types of interactions potentially contributing to the allosteric effect: 1) direct interactions, 2) structural changes and 3) changes in thermal fluctuations. Electrostatic interactions have been implicitly included into the direct interactions. However, they can affect binding at large distances, determined by the Debye screening length. The Debye screening length defines a distance beyond which a charge is effectively screened in the salt solution of concentration c. The Debye length,

defined as

$$\lambda_D = \sqrt{\frac{2\mathrm{e}^2 c}{\varepsilon k_B T}}, \tag{2.36}$$

where e is the elementary charge and ε the absolute permittivity, is approximately equal to 1 nm in physiological salt concentration of 0.1 M. The permittivity of a protein is however substantially lower (≈ 3 [124]) than that of water (≈ 80) resulting of a Debye length of 5 nm.

In this thesis we investigate closely two experimental systems: the dimeric protein CAP and a DNA oligomer binding a drug Hoechst 33258. In both cases the ligands are separated by approximately 1 nm and electrostatic interactions thus cannot be excluded. The focus of this thesis is nevertheless on dynamic allostery and the possibility of electrostatic interaction is discussed only briefly for each system.

Chapter 3

Dynamic Allostery in Oligomeric Proteins

Allostery was for a long time believed to occur purely in oligomeric proteins. Although Volkman showed in 2001 that monomeric proteins can display allostery too [35], the vast majority of known allosteric systems are oligomeric [125]. The ultimate target of us researchers, interested in allosteric effects, is a general theory of cooperative binding. A systematic investigation of individual groups or classes of allosteric systems provides building blocks for such a theory. A study of oligomeric proteins naturally begins with the simplest representatives, proteins with two identical units or homodimers.

The potential contribution of thermal fluctuations to allosteric effects in dimers was first theoretically investigated by Hawkins and McLeish [38]. They focused on heterotropic allostery in dimers consisting of relatively rigid monomers with binding sites located at the dimer interface. The binding was assumed to affect only the interaction between monomers.

Another large group of dimers displays homotropic allostery with the identical binding sites located inside each monomer [125]. Binding to these sites is likely to affect the rigidity of the monomer itself, an effect unaccounted for by the previous model of Hawkins and McLeish [38]. In this chapter we construct a coarse-grained model of a symmetric dimer such that each monomer is assigned one or more internal degrees of freedom and the monomers are elastically coupled. In the first version this model includes only global, low frequency modes. Fast modes may however become coupled to

the slow modes and become involved in the signaling [42]. In order to find possible implications of such coupling and to construct a more realistic model we incorporate fast modes into the model. An analysis of the model at different levels of complexity then elucidates the role of motions on different timescales.

In order to illustrate the application of the model we apply it to a challenging test case, the catabolite activator protein (CAP). CAP displays negative cooperativity upon association with two identical ligands. The conformation of CAP is not affected by the binding, but its vibrational spectrum undergoes a strong modification. Intriguingly the first binding enhances thermal fluctuations yet the second quenches them. We show that this counter-intuitive behaviour is in fact necessary for an optimal anti-cooperative system, and captured within a well-defined region of the model's parameter space. From analyzing the experimental results we conclude that fast local modes take an active part in the allostery of CAP, coupled to the more global slow modes. By including them into the model we elucidate the role of the modes on different timescales. We conclude that such dynamic control of allostery in homodimers may be a general phenomenon and that our model framework can be used for extended interpretation of thermodynamic parameters in other systems.

In the second part of this chapter we demonstrate that the homodimer model can be extended to account for proteins composed of more than two subunits. We first focus on a tetramer (protein composed of four subunits) and exemplify the model on a test case, human haemoglobin. We then outline the generalization of the approach to an oligomer composed of N subunits.

Major part of the work on homodimers presented in this chapter has been published in [126].

3.1 Dynamic Allostery in Homodimers

Proteins with two-fold symmetry constitute a large and important group of proteins. Many DNA-binding proteins, antibodies and receptors are either present in the cell as dimers or are composed of two identical domains [125]. Homotropic allostery has been observed in many homodimers [127–131].

Additionally, it is common in nature that two identical ligands are required to bind to a dimer in order to activate or inhibit its function. Typically only the activation or inhibition is subjected to experimental study, due to its greater biological relevance. The first step of this process, i.e. the binding of the two identical effectors is however likely to be allosteric too, at least in some cases. Popovych *et al.*, investigated one such system: association of two identical effectors called cAMP to a homodimer known as catabolic activator protein (CAP) and showed that the binding is strongly negatively cooperative. This system constitutes the main model system of this chapter and is detailed in Sec. 3.3. Other homodimers are activated or inhibited in similar fashion and we believe that the two identical effectors commonly bind in a cooperative way. Examples of such dimers include trp repressor associating with L-Trp (see Fig. 3.1) [132], PYL protein binding abscisic acid [133] and chorismate mutase binding tryptophan or tyrosine [134, 135].

Because allosteric homodimers are usually large molecules a wide variety of experimental techniques have been used to obtain the experimental evidence. For example, Eaton and Stewart used fluorescence monitored titration to investigate cooperative binding of ATP to a protein histidine kinase CheA homodimer [130], the groups of Kalodimos [4] and of Zuiderweg [129] employed NMR to study negative allostery of the CAP and GCT dimers.

(a)   (b)
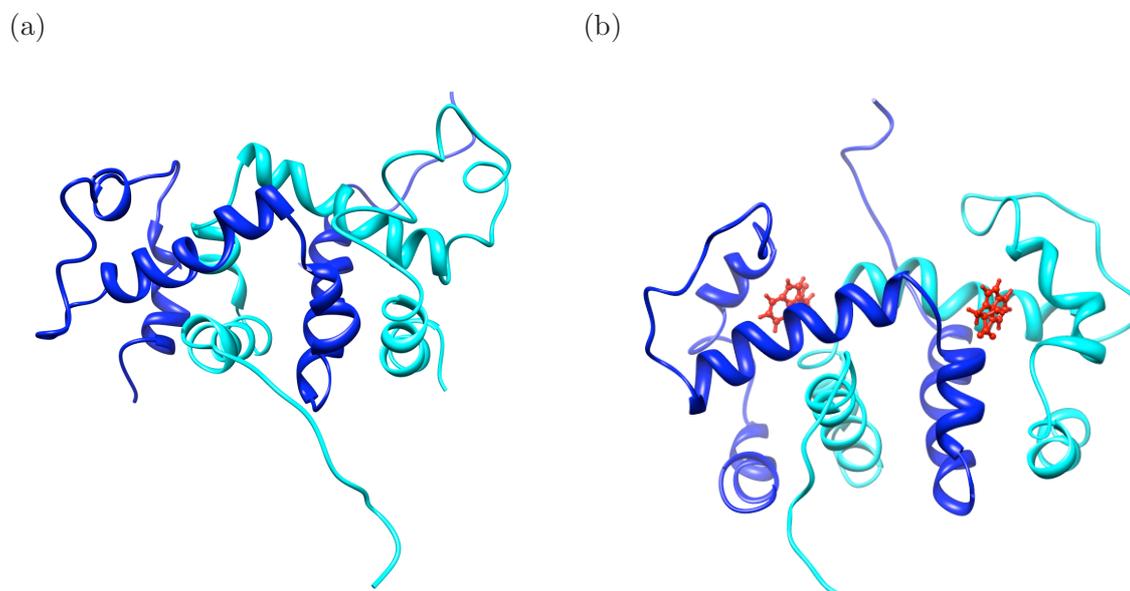
FIGURE 3.1: Three dimensional crystal structure of (a) effector free (pdb-id: 1WRT) and (b) doubly bound (pdb-id: 1WRS) trp repressor. The two monomers of the repressor are shown in different shades of blue, the ligand Trp-L in red.
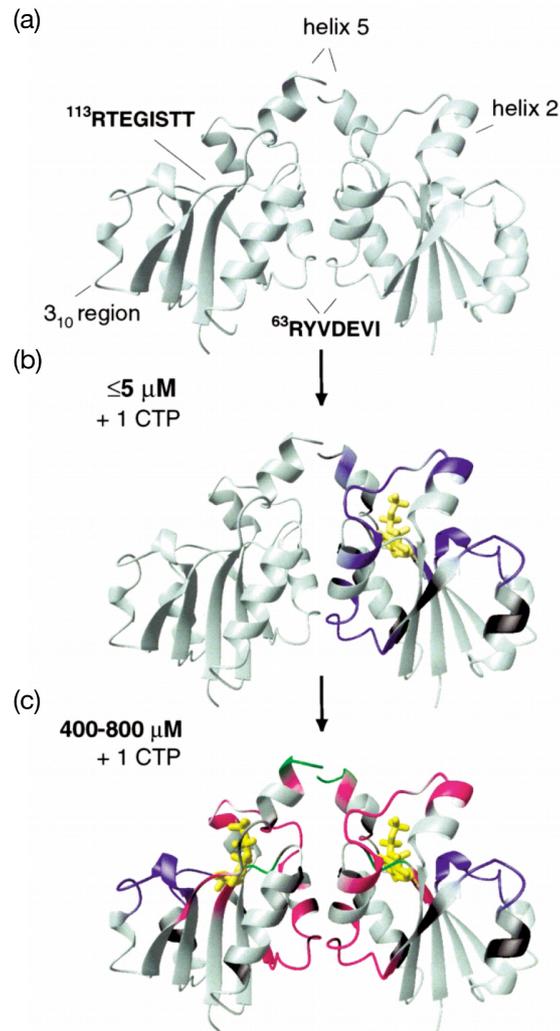
FIGURE 3.2: NMR chemical shifts mapped onto the three dimensional structure of GCT during sequential CTP binding. The shifts demonstrate that the allostery occurs without a significant conformational change. The ligand is shown as a balls-and-sticks model and colored yellow. (b) Distinct changes in chemical shift with respect to the ligand free GCT occur only in the vicinity of the bound ligand and are depicted in blue. Residues colored black are unassigned. (c) The second ligand binding results in more significant change in chemical shift in particular at the dimer interface. Taken from [129].

respectively, various groups have used x-ray crystallography to probe cooperativity in the aspartate receptor [128] and a mixture of techniques such as equilibrium dialysis, gel permeation chromatography, and kinetic enzyme assays was used to measure negatively cooperative binding of several folate substrates to thymidylate synthase dimer [136].

Some experiments indicate that the effectors influence flexibility of the dimer [135] or directly participate in allostery [4, 129]. For example, Stevens *et al.* suggested involvement of dynamics in negatively cooperative binding of a small ligand CTP to the enzyme glycerol-3-phosphate:CTP transferase (GCT) [129]. In their study the NMR spectra revealed that the first binding has negligible influence on the structure of the unliganded binding site (Fig. 3.2). Additionally the authors demonstrated that there is negligible electrostatic interaction between the two negatively charged ligands. Namely, the

titration of the Na⁺ ion into the CTP-saturated protein solution resulted in unmodified NMR spectra. The suggestion that protein dynamics may play a certain role appeared amongst the speculations about the allosteric mechanism. The NMR dynamic analysis was somewhat hampered by limited stability of the unliganded protein but nevertheless yielded interesting results. The first binding step could be mapped only in a sporadic way but seemed to have little impact on the protein dynamics. The second step was however found to be accompanied by a significant rigidification of the whole structure. The authors did leave an open discussion for the allostery mechanism. Several years later the same pattern of dynamic changes was observed in CAP and was concluded to give rise to entropically driven allostery [4].

Catabolite activator protein (CAP) has been shown to display negative cooperativity without a significant conformational change upon binding two identical ligands called cAMP [4]. NMR measurements revealed that CAP's fluctuations undergo a counter-intuitive change upon binding, whereby binding of the first cAMP molecule slightly enhanced, and the second completely suppressed the amplitude of global motions. Due to much larger amount of data available on this protein we select CAP over GCT as our model system.

3.2 Model of a Homodimer

As mentioned earlier, a protein undergoes thermal fluctuations with frequencies spanning several orders of magnitude. Out of these the lowest frequency modes are mainly responsible for communication of the allosteric signal across large distances whilst fast modes are localized and can only amplify or quench the signal. Hence we start by constructing a model of slow modes only and analyse the dynamical parameter space for configurations yielding positive and negative allostery. In this model we first assign a single internal breathing mode to each monomer and elastically couple the monomers. Later we increase the number of internal modes and add the fast modes to reflect the complexity of real homodimers and to obtain biologically relevant values of allosteric free energy. An important assumption is that the monomeric effective spring stiffnesses can only be altered locally, and not distantly, by a ligand binding. Fast modes are added in the same way as described in Sec. 2.5.

3.2.1 Single Slow Mode

A homodimer in this context is a protein consisting of two identical subunits, each of which binds a ligand. In the first and simplest approximation of the equilibrium dynamics of such protein we assign one internal, "breathing", mode to each subunit and then elastically couple the subunits. This very simple and coarse-grained model is designed to explore only the qualitative features of dynamic allostery in the system. For the unliganded protein the internal mode and coupling strength are characterized by spring constants k and k_c respectively. Binding of a ligand is modeled as changes of the spring constants.

We concentrate on the symmetric case where the two ligands and their binding sites are identical. We make two assumptions on the effect of the ligand binding. The first follows from the symmetry of the system and requires that both binding events have the same effect on the spring constant representing the protein. In the second we assume locality: the effect of binding is not directly propagated to the distant subunit. At this level of model, locality means that ligand binding to one subunit affects only the stiffness of its own internal mode and the coupling to the other subunit, but no direct effect on the internal stiffness of the other subunit. The ligand binding alters chemical bond structure locally and therefore only the spring constants that directly derive from these bonds are likely to change. However the subunits are elastically coupled and thus the thermal motions of the distant subunit are indirectly modified too, leading to the dynamic allosteric effect. The assumptions are demonstrated in Fig. 3.3. We define non-dimensional parameters describing the effect of substrate binding as follows: the first binding event changes the local subunit spring constant by a factor β and the coupling spring constant by a factor α. Introduction of the second ligand evokes the same alteration in the other subunit.

The system is mathematically described by a Hamiltonian

$$\mathcal{H} = \frac{1}{2}\mathbf{p^T}\mathbb{M}^{-1}\mathbf{p} + \frac{1}{2}\mathbf{x^T}\mathbb{K}\mathbf{x}. \tag{3.1}$$

The inertia matrix $\mathbb{M}$ is approximately constant during the binding events and therefore can be left out from the subsequent calculations. For the unliganded protein the elastic
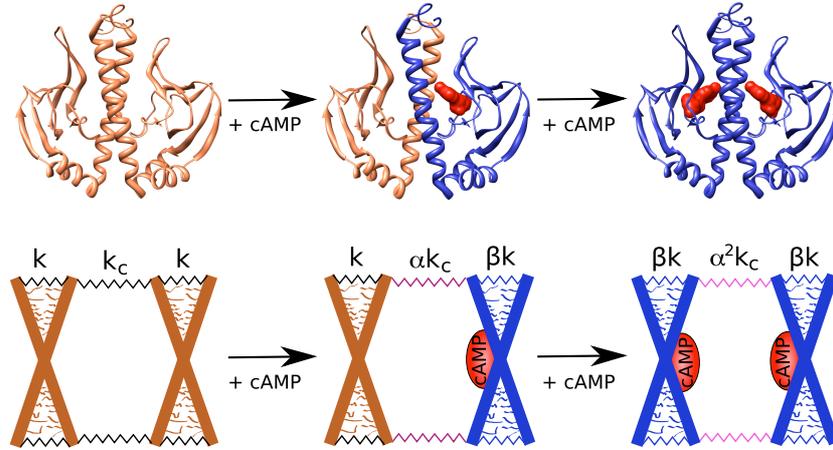
FIGURE 3.3: Residues 1-138 of crystallographic structure of CAP (PDB entry 1G6N) binding the ligand cAMP (*red*) and a sketch of the corresponding coarse-grained model of the system. The large X represents the backbone of one subunit whose one internal mode is simulated by a scissor like movement of the rods. The little protrusions represent small structures moving fast relative to the slow scissor like motion of the rods. The internal mode of each subunit and the coupling is defined by the elastic constant k and k_c respectively. The constants are altered upon binding by factors α and β

part of the Hamiltonian reads

$$\frac{1}{2}\mathbf{x^T}\mathbb{K}\mathbf{x} = \frac{1}{2}\left(x_1\ x_2\right)\begin{pmatrix} k+k_c & -k_c \\ -k_c & k+k_c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \tag{3.2}$$

where x_1 and x_2 are the generalised amplitudes of the internal modes in each of the individual subunits. The partition function of the coarse-grained dimer undergoing structural fluctuations is obtained from the Hamiltonian and reads

$$Z = \left(2\pi k_B T\right)^2 \left(\frac{|\mathbb{M}|}{|\mathbb{K}|}\right)^{1/2}, \tag{3.3}$$

where k_B is the Boltzman constant. The free energy is then, $G = -k_B T \ln Z$. We are only interested in the free energy differences between the ligation states and therefore all terms that stay constant during the binding can be ignored. We wish to calculate only the dynamic contributions and therefore other contributions such as entropy of desolvation or hydrophobicity of the binding pockets are not included in this calculation.

The requirements for the two constraints of symmetry and locality of binding are implemented by introducing coefficients α and β into the matrix $\mathbb{K}$ as illustrated in Fig. 3.3. The singly bound protein is thus described by a Hamiltonian with the elastic part equal

to

$$\frac{1}{2}\mathbf{x^T}\mathbb{K}\mathbf{x} = \frac{1}{2}\left(x_1\ x_2\right)\begin{pmatrix} \beta k + \alpha k_c & -\alpha k_c \\ -\alpha k_c & k + \alpha k_c \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{3.4}$$

The Hamitonian of the doubly liganded protein is found analogously.

The difference between the free energy change of each binding step ($\Delta\Delta G$) measures the degree of cooperativity (see Sec. 2.5); $\Delta\Delta G = (G_{2:1} - G_{1:1}) - (G_{1:1} - G_{APO})$, 2:1 refers to doubly liganded and 1:1 to singly liganded protein. Let us recall the sign convention, $\Delta\Delta G \neq 0$ indicates cooperativity, $\Delta\Delta G < 0$ corresponds to a positively and $\Delta\Delta G > 0$ to negatively cooperative system. A larger absolute value of $\Delta\Delta G$ signifies a more cooperative system. The evaluation yields

$$\Delta\Delta G = \frac{1}{2}k_B T \ln\left(\frac{(\beta^2 + 2\alpha^2\beta K_c)(1 + 2K_c)}{(\beta + \alpha K_c + \alpha\beta K_c)^2}\right), \tag{3.5}$$

where the crucial quantity $\Delta\Delta G$ is now expressed using three dimensionless parameters: $K_c = k_c/k$ the ratio between the subunit and the coupling spring constant, α, the dimensionless enhancement of the coupling strength on binding and β the dimensionless enhancement on the local subunit mode stiffness (Fig. 3.3). The dimensionless character of the equation is advantageous for the parameterisation from experimental results, because only relative changes in the spring stiffness contribute to $\Delta\Delta G$. We look for areas in the parameter space yielding $\Delta\Delta G \neq 0$. To picture the three parameter space we make two fixed choices in each of two qualitatively regimes for the parameter α and plot $\Delta\Delta G$ as a function of the remaining two parameters (Fig. 3.4 top).

The parameter space is divided into two subspaces: $\alpha > 1$ which corresponds to stiffening of the coupling between subunits on binding of a ligand and $\alpha < 1$ corresponding to coupling loosening. The shape of the $\Delta\Delta G$ landscape is non-trivial for $\alpha \neq 0$; regions of positive and negative cooperativity are observed in both subspaces. The qualitative character of the landscape is independent of the choice of the value of α within each subspace, however there are substantial differences between the two subspaces (Fig. 3.4). In the case where coupling stiffens, positive or negative cooperativity is achieved by carefully choosing β; if the coupling loosens, K_c becomes the critical parameter instead. The second major difference is that as α tends to 0, $\Delta\Delta G$ becomes more positive. When $\alpha > 1$ the values in the area where $\Delta\Delta G > 0$ are slightly enhanced, however for larger values of β the landscapes cross and the system becomes increasingly cooperative (Fig. 3.4). This suggests that positively cooperative systems are exploring the subspace
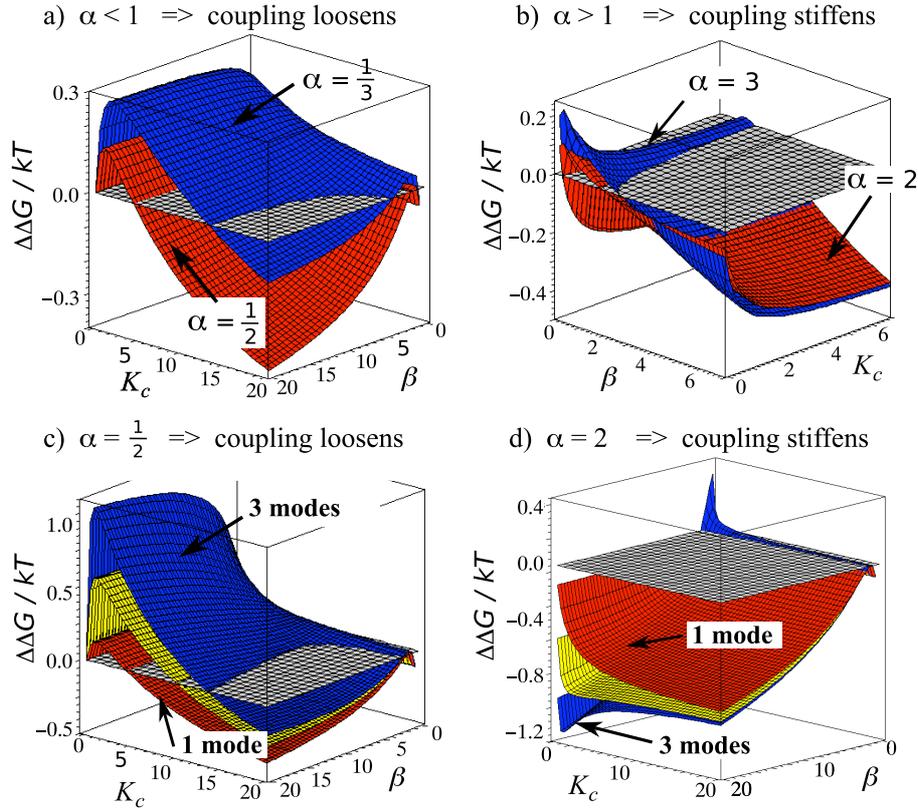
FIGURE 3.4: Top: allosteric free energy landscapes for a single slow mode. Bottom: allosteric free energy landscape for one (*blue*), two (*yellow*) and three (*red*) slow modes. The plane $\Delta\Delta G = 0$ is shown to highlight the border between positive ($\Delta\Delta G < 0$) and negative cooperativity ($\Delta\Delta G > 0$).

$\alpha > 1$ and negatively allosteric system the subspace $\alpha < 1$. The borderline case of $\alpha = 1$ does not result in negative cooperativity for any choice of the remaining parameters.

The allosteric free energy (Eq. 3.5) is directly proportional to the temperature implying (see Eq. (2.24)) that the slow mode change gives rise to purely entropic allostery (in the isothermal case).

A good measure of slow mode amplitudes is provided by the mean relative fluctuations $\langle (x_1 - x_2)^2 \rangle$

$$\langle (x_1 - x_2)^2 \rangle = \frac{1}{Z} \iint dx_1 \, dx_2 \, (x_1 - x_2)^2 \exp\left(-\frac{\mathcal{H}(x_1, x_2)}{k_B T} \right). \tag{3.6}$$

This mean is evaluated for each ligation state and yields

$$\left\langle (x_1 - x_2)^2 \right\rangle_{free} = \frac{1}{2} \frac{k_c}{k(k + 2k_c)}$$

$$\left\langle (x_1 - x_2)^2 \right\rangle_{1:1} = \frac{1}{2} \frac{\alpha k_c}{k(\beta k + \alpha k_c + \alpha \beta k_c)}$$

$$\left\langle (x_1 - x_2)^2 \right\rangle_{2:1} = \frac{1}{2} \frac{\alpha^2 k_c}{\beta k(\beta k 2\alpha^2 k_c)}. \tag{3.7}$$

The results are stated in terms of the original spring constants.

The ratios of the mean relative fluctuations are dimensionless and are used to evaluate the evolution of the fluctuations during the sequential binding. Four types of behaviour are observed: 1) and 2) sequential increase and decrease of fluctuation amplitude respectively, 3) fluctuations are amplified upon the first ligand binding but quenched upon the second binding. The fluctuation amplitude of the doubly liganded state is smaller than that of the unliganded protein, 4) increase in the amplitude is followed by decrease, however the fluctuation amplitude of the doubly liganded state is now larger than that of the unliganded system. The four types of behviour are mapped onto the allosteric free energy surface in Fig. 3.5. We observe that for $\alpha < 1$ all four types of behaviour are present for large regions of the parameter space. The most interesting observation however is that in order to maximize negative cooperativity ($\Delta\Delta G > 0$) the loosening-tightening effect is required (case 3). In the case of $\alpha > 1$ the fluctuations tend to be sequentially quenched upon the binding, in particular this is the case for a positively cooperative system. However a negatively cooperative system whose coupling would get stronger upon binding is again likely to display the loosening-tightening effect.

The monomeric average fluctuations $\left\langle x_i^2 \right\rangle$, $i = 1, 2$ are evaluated analogously to the relative fluctuations Eq. (3.6). The resulting landscapes are relatively complex but several general observations can be made. Let us focus on the more common case where the ligand binding has a local stiffening effect ($\beta > 1$). If additionally $\alpha > 1$ then the average fluctuations decrease upon each ligand binding. When however $\alpha < 1$ then more complex situations arise, where for some parameter choices motions are activated either in the unliganded monomer or throughout the whole dimer. We discuss the monomeric average fluctuations in detail on the example of CAP in Sec. 3.3.

Even the simplest level of coarse-grained model shows that allosteric effects can arise in coupled dimers purely from spatial fluctuations. The evaluation of the fluctuations
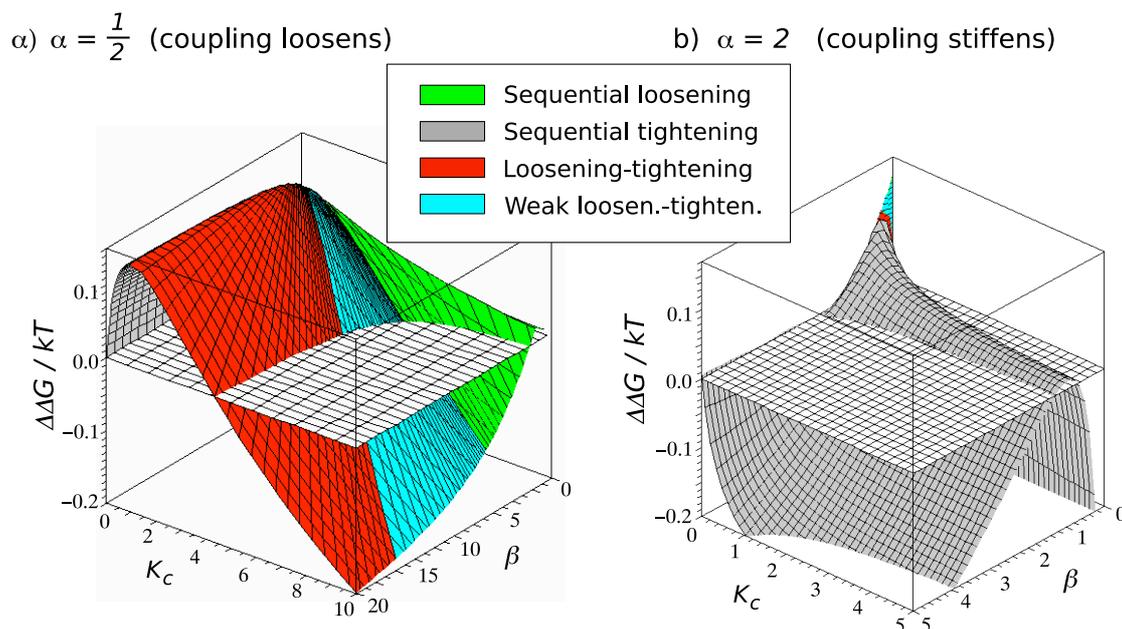
FIGURE 3.5: Four regions with different change in fluctuations mapped onto the $\Delta\Delta G$ landscape for $\alpha = 1/2$ (left) and $\alpha = 2$ (right). Color code: in the red region the loosening-tightening effect is observed. The fluctuations of the doubly liganded system are smaller than that of the apo-protein. The blue region is characterized by the weak loosening-tightening effect, whereby the doubly liganded system moves more than the apo-protein, but less than the 1:1 system. The green region is characterized by sequential stiffening of the protein upon binding. In the grey region each binding increases the fluctuations. The green region for $\alpha > 1$ is hidden behind the peak in this view.

demonstrates that the loosening-tightening effect is required to produce strong negative cooperativity, whereas strong positive cooperativity is accompanied by sequential tightening of the system. However the allosteric free energy is of purely entropic origin and its values of $\Delta\Delta G$ match the generally observed values of few $k_B T$ only for limiting cases of parameters tending to 0 or ∞. That represents unphysical conditions and we conclude, as might be expected on physical grounds, that more modes naturally present in the system must take part in the allosteric signaling. These fall into two classes: fast local modes and additional, global, slow modes.

The effect of fast modes on the allostery have been investigated before [42] and it has been shown that the net values of $\Delta\Delta G$ are not amplified but that the free energy is split into compensating entropic and enthalpic part, which themselves do acquire enhanced absolute values. The effect of including fast modes in the model of a homodimer will be discussed at the end of this section.

3.2.2 Multiple Slow Modes

We extend our model to include M slow modes per subunit. We assume that the modes are harmonically coupled to each other across the subunits. This corresponds to a Hamiltonian

$$\mathcal{H} = \sum_{i=1}^{2}\sum_{j=1}^{M} k_{i,j} x_{i,j}^2 + \sum_{i,k=1}^{2}\sum_{j<l}^{M} \lambda_{[i,j][k,l]}(x_{i,j} - x_{k,l})^2, \tag{3.8}$$

where $x_{i,j}$ is a coordinate of j-th mode on the i-th subunit with the respective spring stiffness $k_{i,j}$. The coupling constants $\lambda_{[i,j][k,l]}$ are in principle different for all modes and can be parameterised from experiments or simulations. At this level in order to probe the properties of the model, while avoiding a proliferation of arbitrary parameters, we constrain their value by reasonable simplifying assumptions. We set all coupling and internal subunit constants equal to each other for the free symmetric protein, i.e. $\lambda_{[i,j][k,l]} = k_c$, and $k_{i,j} = k$, $\forall\, i,j,k,l$. The Hamiltonian reduces to

$$\mathcal{H} = \sum_{i=1}^{2}\sum_{j=1}^{M} k\, x_{i,j}^2 + \sum_{i,k=1}^{2}\sum_{j<l}^{M} k_c(x_{i,j} - x_{k,l})^2. \tag{3.9}$$

We further assume that as in the one mode case: (i) ligand binding affects only the local elastic constants plus the coupling constants, (ii) binding of the ligand to either subunit has the same effect, all internal subunit stiffnesses change by a factor β and coupling constants by a factor α. The resulting Hamiltonian of the apo-protein in matrix form is $\mathcal{H} = \mathbf{x^T}\mathbb{K}_0\mathbf{x}$ where

$$\mathbb{K}_{0_{i,j}} = (k + 2Mk_c)\delta_{i,j} - k_c, \quad i,j = 1,\dots,2M, \tag{3.10}$$

and δ denotes the Kronecker delta. The matrix $\mathbb{K}_1$ of the singly liganded complex has alternating terms $\beta k + \alpha(2M - 1)k_c$ and $k + \alpha(2M - 1)k_c$ on the diagonal and the off-diagonal terms are equal to $-\alpha k_c$. The diagonal terms of the matrix $\mathbb{K}_2$ of the doubly liganded complex are $\beta k + \alpha^2(2M - 1)k_c$, off-diagonal $-\alpha^2 k_c$. The allosteric free energy is obtained from the partition function as previously described,

$$\Delta\Delta G = \frac{1}{2} k_B T \ln\left(\frac{|\mathbb{K}_0|\,|\mathbb{K}_2|}{|\mathbb{K}_1|^2} \right), \tag{3.11}$$

and yields

$$\Delta\Delta G = \frac{1}{2} k_B T \ln\left[\frac{\beta\left(\beta + 2M\alpha^2 K_c\right)^{2M-1}\left(1 + 2MK_c\right)^{2M-1}}{\left(\beta + 2M\alpha K_c\right)^{2M-2}\left(1 + 2M\alpha K_c\right)^{2M-2}\left(\beta + M\alpha\beta K_c + M\alpha K_c\right)^2}\right].$$
(3.12)

We observe that $\Delta\Delta G$ is a function of four dimensionless parameters α, β, K_c and number of modes M. The central result is that the free energies are indeed modified with increasing number of slow modes as is shown in Fig. 3.4.

In particular, this extension confirms that negatively allosteric systems are likely to live in the $\alpha < 1$ subspace and positively cooperative in $\alpha > 1$ subspace. In these subspaces including extra slow modes leads to the amplified allosteric effect in question. In the subspace $\alpha > 1$ this amplification is observed also in the region with $\Delta\Delta G > 0$ but is much less pronounced than in the other subspace. $\Delta\Delta G$ values of ± 5 $k_B T$ are observed for as few as 5-10 slow modes. The values in connection to experiments will be discussed in more detail in section 3.3.

The fluctuation changes are evaluated in the form of the fluctuation matrix $\mathbb{C}$

$$\mathbb{C}_{ij} = \left\langle (x_{1,i} - x_{2,j})^2 \right\rangle.$$
(3.13)

As more slow modes are added to the system, the fluctuation amplitude per mode $\frac{1}{M^2}\sum_{i,j=1}^{M}\mathbb{C}_{ij}$ decreases while the total fluctuations $\sum_{i,j=1}^{M}\mathbb{C}_{ij}$ increase (Fig. 3.6). The comparison of the fluctuations of three ligation states yields again the same four types of behavior as the simple-mode case depending on the parameter choice (Fig. 3.5). This is observed for any number of modes M. The mapping onto the allosteric free energy landscape results in an analogous picture to the simple-mode case too, the four classes of behaviour span the same regions of the $\Delta\Delta G$ landscape.

3.2.3   Fast modes

In contrast to the slowest modes, fast modes are typically localized (involve only a few atoms) and are therefore unlikely to transmit allosteric signal across large distances by themselves [120]. However they can couple to the slow modes and so become involved in the transmission, modifying its amplitude. Here we draw on previous work [42] to couple several fast modes to the global, slow ones in the way explained in the Methodology
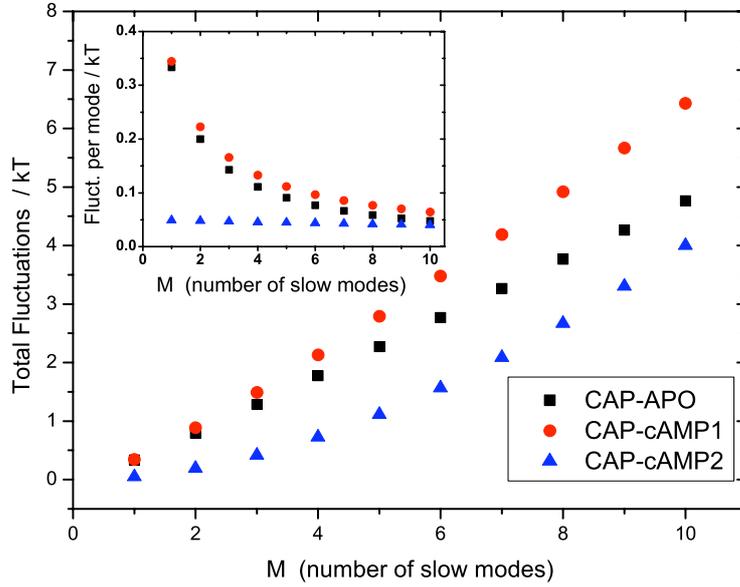
FIGURE 3.6: Slow mode fluctuations as a function of number of slow modes M for chosen parameters $K_c = 1$ and $\beta = 20$. Fluctuations per mode decrease whereas the total fluctuations increase with number of slow modes included in the model. The loosening-tightening effect becomes stronger with M.

section 2.5. We can picture the situation as shown in Fig. 3.3. The slow breathing mode of the subunit is represented as a scissor-like movement of the two rods. Fast motions of smaller structures within the subunit such as side chains are represented as vibrations of little protrusions attached to the rods. Here we derive the results for one slow and multiple fast modes and then generalize the result for multiple slow and fast modes in the next section.

The principle of coupling was explained in Sec. 2.5. It is based on the idea that the flexibility of the fast modes increases with the amplitude of the slow mode. We assume therefore that the rigidity of the fast mode depends on the absolute displacement $|x_s|$ where x_s stands for x_1 or x_2, the coordinate of the larger structure within the slow mode of the individual monomer. A further physically-motivated assumption made is that the fast modes are only coupled to the local slow mode (Fig. 3.3) and that the fast mode potential becomes shallower and flatter with the increased $|x_s|$.

We implement the idea by modifying the Hamiltonian to

$$\mathcal{H} = \mathcal{H}_s + \sum_{i=1}^{2N} V_{f_i}, \tag{3.14}$$

where $\mathcal{H}_s$ is the Hamiltonian of the slow modes (Eq. 3.2) and the sum adds up the fast modes. N fast modes are enslaved to each subunit, the i-th enslaved mode experiences a potential V_{f_i}. If the fast mode is not coupled to the slow mode its effective potential is the harmonic approximation

$$V_{f_i} = -V_{f0} + \frac{k_f x_{f_i}^2}{2}, \tag{3.15}$$

the potential depth V_{f0} and the mode stiffness k_f are assumed as in the previous work [42] to be the same for all fast modes. The width and depth of the potential are assumed to be affected by the slow mode in the coupled case. The increased flexibility of the slow mode corresponds to a flatter and wider potential V_{f_i} for which we take the functional parametrization

$$V_{f_i} = -V_{f0}\left(-\frac{k_v x_s^2}{k_B T} + 1\right) + \frac{1}{2}\left(\frac{k_f}{\exp\left(\frac{k_k x_s^2}{2 k_B T}\right)}\right) x_{f_i}^2, \tag{3.16}$$

$x_s = x_1, x_2$ is the slow mode coordinate, x_{f_i} the i-th fast mode coordinate, k_v, k_f and k_k are coupling constants (given, without loss of generality, the dimensions of a spring constant). The choice of coupling functions is arbitrary, the only requirement is smooth widening and flattening of the potential with increasing $|x_s|$. We select these functions over those used by Hawkins and McLeish [42] because no approximation is required to obtain the partition function. We repeat the statistical mechanics calculation with the modified Hamiltonian and find

$$\Delta\Delta G = \frac{1}{2} k_B T \ln \frac{[(\beta + \alpha^2 K_c + AN)^2 - \alpha^4 K_c^2][(1 + K_c + AN)^2 - K_c^2]}{[(\beta + \alpha K_c + AN)(1 + \alpha K_c + AN) - \alpha^2 K_c^2]^2}, \tag{3.17}$$

where

$$A = \frac{V_{f0} k_v}{k k_B T} - \frac{k_k}{4k}. \tag{3.18}$$

The parameters α, β and K_c define the slow mode during the two binding steps (see Fig. 3.3).

Including fast modes has a relatively complex effect on the values of allosteric free energy. In particular if $A < 0$, i.e. when the coupling of the slow potential width is significantly

stronger than coupling to the potential depth, the situation is difficult to generalize. For some special choices of the parameters the maximum of $\Delta\Delta G$ is slightly increased but generally the values of $\Delta\Delta G$ are smaller than in non-enslaved case.

If $A > 0$ then the region of the parameter space α, β, K_c yielding $\Delta\Delta G > 0$ increase. However the absolute maximal value of $\Delta\Delta G$ is always slightly lower than in the non-enslaved case. The structure of the coupled model is most clearly seen if we make the simplifying choice of $A = 0$ which would correspond to a system at a fixed special temperature. Now $\Delta\Delta G$ is identical to the non-enslaved case. However the free energy is now composed of compensating entropic and enthalpic terms. This enthalpic contribution arises naturally from the coupling between the mean energy adopted within the local mode potentials and the amplitude of the global dynamics, and can be calculated by standard application of thermodynamic relations to our model. For isothermic changes

$$H = k_B T^2 \frac{\partial \ln Z}{\partial T} \tag{3.19}$$

and thus

$$\Delta\Delta H = N V_{f_0} K_v \left(\frac{\beta + \alpha^2 K_c}{\beta^2 + 2\alpha^2 \beta K_c} - \frac{1 + \beta + 2\alpha K_c}{\beta + \alpha\beta K_c + \alpha K_c} + \frac{1 + K_c}{1 + 2K_c} \right). \tag{3.20}$$

The allosteric enthalpy can be positive or negative, however for $\alpha > 1$ the negative values of $\Delta\Delta H$ occur only in a very small region of the parameter space and are of very small absolute value. In the case of $\alpha < 1$ on the other hand this region and the values become significantly larger. In order to compare the values of enthalpy to the values of entropy and find the region of parameter space in which the enthalpy entropy compensation is taking place we would need to know the values of coupling parameters and the fast potential depth.

3.3   An Example: Catabolite Activator Protein

To illustrate the utility of our model we apply it to an example homodimer, the catabolite activator protein (CAP). This transcriptional activator in *E. Coli*, consists of two identical subunits each of which binds a small activator called cAMP (cyclic adenosine monophosphate). The cAMP molecules serve as an allosteric activator that greatly increases the CAPs affinity for DNA. The binding of the two cAMP molecules to the

protein is itself allosteric and negatively cooperative; the binding of the first cAMP molecule reduces the affinity for the second by nearly two orders of magnitude [137]. The distance between the two cAMP ligands is $10\,\text{Å}$ [138] excluding electrostatic or any other direct interactions.

Each subunit of CAP is composed of two distinct domains, the cAMP binding domain (residues 1-138) and the DNA binding domain (residues 139-209). The negative cooperativity upon cAMP binding takes place independently of the presence of the DNA binding domain and according to Heyduk *et al.* becomes even stronger in its absence [139]. Popovych *et al.* studied the allosteric binding of cAMP in the truncated version of CAP (CAP^N, residues 1-138) [4]. They recorded NMR spectra of all three ligation states (Apo-CAP^N, cAMP$_1$-CAP^N and cAMP$_2$-CAP^N) and extracted information on the protein structure and dynamics.

Structural changes were assessed from chemical shift mapping. Chemical shifts are extremely sensitive to the environment (see Sec. 1.4.1 for explanation) and thus report on very subtle changes in the average protein structure. Figure 3.7 shows the chemical shift changes during the individual binding steps mapped onto the structure of the protein. We observe that each binding results in a pronounced change in the average structure of the interacting monomer but the other monomer's structure remains practically intact.

The dynamics of the system was probed across a wide range of timescales. The slow,
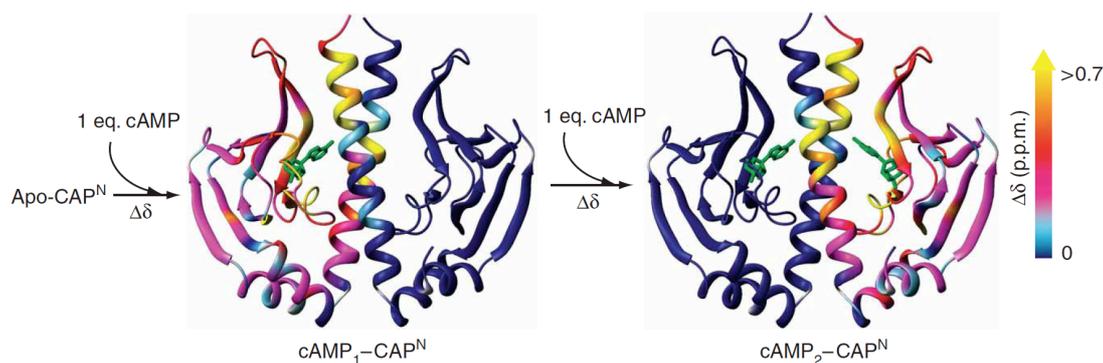
FIGURE 3.7: Change in chemical shifts during the two cAMP binding steps mapped onto the structure of CAP^N reports on locally modified protein conformation. Large values of the shift (in yellow) indicate large changes in structure, zero values signify region intact by the binding. The cAMP ligand is shown as green sticks. Taken from [4].

Apo-CAP^N cAMP$_1$-CAP^N cAMP$_2$-CAP^N
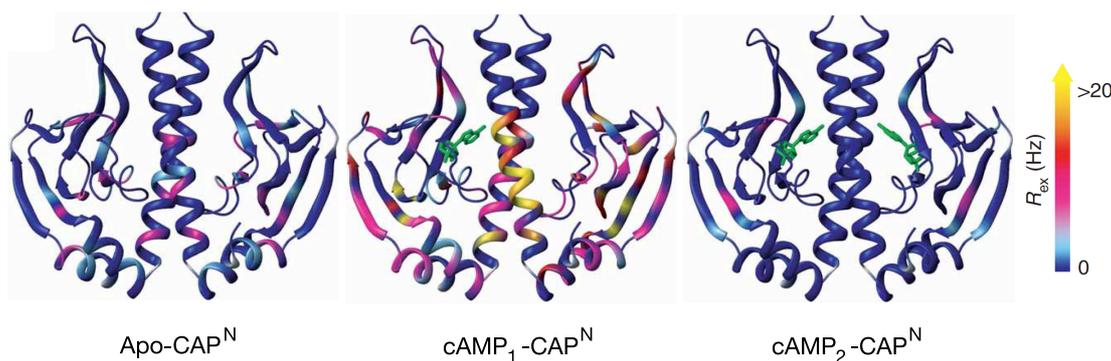
FIGURE 3.8: Effect of cAMP binding on the slow (μs-ms) motions of CAP^N mapped onto the structure of the protein. The color scale ranges from dark blue corresponding to rigid areas to yellow corresponding to flexible residues. The ligand cAMP is shown as green sticks. Taken from [4].

μs-ms motions were obtained by relaxation measurements: both T_2-CPMG and rotating frame data were recorded. The excess transverse relaxation rate R_{ex} was extracted from both experiments. R_{ex} denotes the contribution to the relaxation time T_2 arising from fluctuations between different conformations (states with a different chemical shift). The first cAMP binding results in a pronounced and widely spread increase in R_{ex} signifying activation of the slow motions. Notably, the motions are activated in both subunits. The second cAMP binding, on the other hand, completely quenches the μs-ms motions. The change in R_{ex} upon the individual binding is shown in Fig. 3.8. The relaxation dispersion measurements further indicate that large portions of the protein move in a concerted fashion.

The fast backbone motions on ps-ns timescale were characterised in the same work [4] by reduced spectral density mapping (RSDM). Details of this technique can be found e.g. in ref. [140]. Briefly, the spectral density function at a special frequency ($0.87\omega_H$, where ω_H is the ^{1}H angular Larmor frequency) is calculated because it reports effectively on the fluctuations of internuclear vectors. The data from these measurements are shown in Fig. 3.9 and indicate that fast motions are affected substantially less by cAMP binding than the slow modes. They display sequential tightening, but the first binding has less pronounced effect than the second.

In summary, the NMR relaxation measurements of Popovych *et al.* rule out ligand-induced conformational change in the binding site of the second ligand but indicate that a substantial modification in the dynamic behavior takes place. The slow backbone
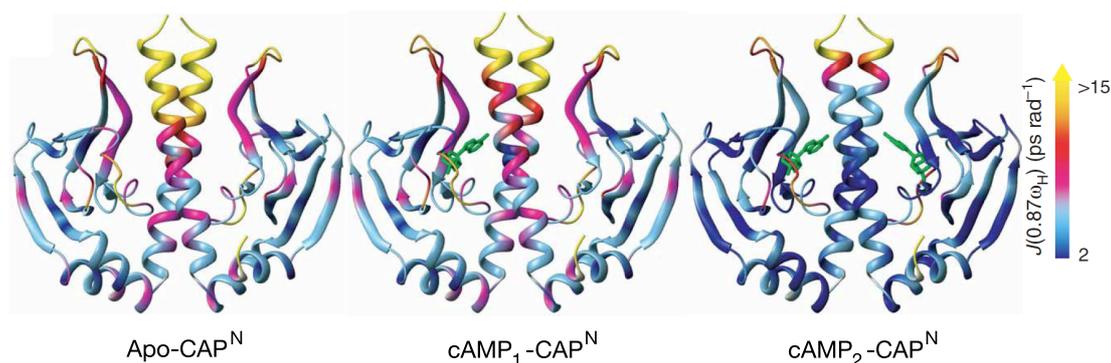
FIGURE 3.9: Effect of cAMP binding on the fast (ns-ps) backbone motions of CAP^N. Values of the spectral density function are mapped onto the structure of the protein with the following colour coding: low values (in blue) mark rigid areas and high values (in yellow) flexible areas. The ligand cAMP is shown as green sticks. Taken from [4].

motions (μs-ms time scale) exhibit a non-intuitive pattern whereby binding of the first cAMP molecule slightly enhances, and the second completely suppresses the amplitude of these global motions. Fast motions of the backbone on the ps-ns timescale change far less than the slow motions.

Thermodynamic potentials of the individual binding steps were obtained by the same group [4] from calorimetric measurements. The measured positive value of the allosteric free energy $\Delta\Delta G$ = $4.7\,k_BT$ confirms negative cooperativity, yet the enthalpic term ($\Delta\Delta H$ = $-1.8\,k_BT$) actually favors binding of the second cAMP ligand. The authors concluded that the strongly unfavorable entropy ($T\Delta\Delta S$ = $-6.5\,k_BT$) drives the negative cooperativity.

In the previous section we derived the structure of the allosteric free energy landscape arising from ligand induced change in slow motions for a coupled dimer. The main assumptions were that the individual ligand bindings have local and identical effect on the slow modes of the protein. In order to check the validity of these assumptions for the case of CAP we used a selection of elastic network models and other freely available programs and analysed the crystal structure of the protein. The programs used were

- **iGNM** - Gaussian network model that uses C_α representation of the protein.

- **ElNemo** - elastic network model that uses rotation translation block (RTB) representation of the protein.

- **WEBnm** - elastic network model that uses C_α representation of the protein.
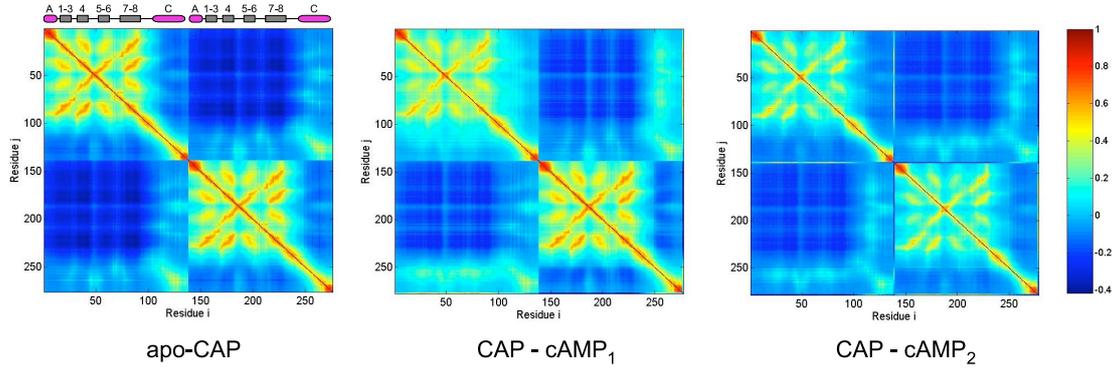
FIGURE 3.10: Cross correlation map, Λ_{ij}, between residue i and j, for three ligation states of CAP^N, obtained from Gaussian network model implemented on the webserver iGNM. A pair subjected to a fully correlated motion ($\Lambda_{ij} = 1$) is colored dark red, fully anti-correlated motions (Λ_{ij}) are not present and moderately correlated motions are colored dark blue. cAMP binding disturbs correlations in the liganded monomer (top left corner of the middle picture) but introduces correlation between the central helices and the liganded monomer. Binding of the second cAMP reestablishes symmetry in the motion pattern and removes correlations of the central helices to the β-sheet structures. Main parts of the secondary structure of CAP are shown above the apo-CAP map, α-helices are represented as magenta cylinders and β-sheets as grey rectangles.

- **FIRST/FRODA** - a module for rigidity analysis of a protein that produces a coarse-grained model of the molecule and then performs a Monte Carlo simulation of the protein dynamics using this model.

These methods are introduced in Sec. 2.3 and 2.4.

These methods appear ideally suited for study of low frequency protein dynamics. In reality problems arise when e.g. effect of small ligand binding on these low frequency modes is studied. The models are coarse-grained in such a way that binding of a small ligand is sometimes not well reflected in the calculated motions. From the comparison of the above programs we concluded that only iGNM is of practical use in the question of allostery of CAP. The detailed results from all programs are shown in Appendix A.

We conducted a systematic study of CAP^N in its three ligation states using the Gaussian network model [108], implemented on the webserver iGNM [112]. All simulations were performed on CAP^N; the atomic model was obtained from the crystal structure of the doubly liganded full length protein (pdb entry 1G6N) by selecting desired residues (1-138) and stripping off cAMP ligands for the singly liganded and unliganded version.

(a)

(b)
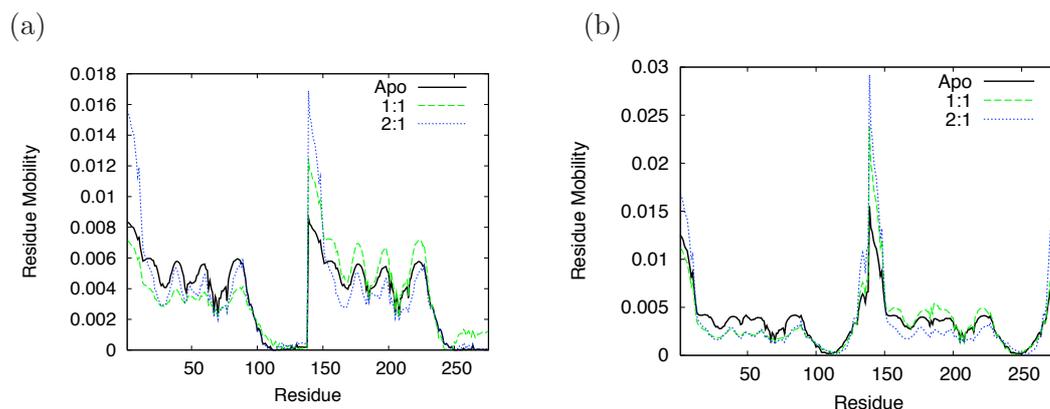
FIGURE 3.11: Residue mobilities (square displacements of the individual residues) induced by (a) 1st mode and (b) average of 1st-3rd mode obtained from the iGNM server. The three ligation states of CAP: apo-CAP (in black), singly liganded CAP (in green) and doubly liganded protein (in blue) are shown. In the singly liganded protein cAMP is associated with the first monomer (residues 1-138), mobilities of cAMP itself are not shown. (a) The prominent peaks at residues 1-7 and 139-145 correspond to the small α helices at the N-terminus. A discrepancy is observed between several iGNM files as some show the fluctuations of these helices substantially smaller. The remaining data agree. (b) Increased mobility is observed also at the C-terminus of each chain (residues 125 -138) corresponding to the random coil at the end of the central helices.

iGNM is a beads-and-springs model that represents every residue as a bead located at the C_α atom. We modified the pdb file so that a bead is also placed in the middle of each cAMP ligand. For more details see Appendix A.

The evolution of the dynamic behaviour is best manifested dynamical cross-correlation maps, $\Lambda_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle$ between residues i and j (Fig. 3.10). We observe that the two main subunits of CAP are very little correlated in the apo-protein which implies that subunits move as weakly coupled individual units. cAMP binding strengthens correlation between the central helices and the β-sheet structure of the liganded monomer, confirming that communication between the two subunits does not proceed directly but only through the interface (central helices). The dynamical pattern of the unliganded subunit is approximately unperturbed, which also motivates the assumption we make on the coarse-grained effect of coupling.

The dynamical changes can be also well observed from the mobilities (square displacements) of the individual residues induced by the low frequency modes. The displacements induced by the 1st mode and those averaged over lowest three modes are plotted in Fig. 3.11. In Appendix A these mobilities are also shown mapped onto the structure of CAP^N and confirm that each subunit consists of two rigid domains (the central helix

and the β-sheet structure) that move with respect to each other. The most interesting observation is made on the effect of ligand binding. The first cAMP binding introduces an obvious asymmetry to the dynamic pattern of CAP. The liganded monomer becomes much stiffer, the ligand provides an anchor between the previously loosely attached β-sheet structure and the central helix. The unliganded monomer's motions on the other hand become activated upon this binding. The entropic penalty for suppressing the amplified motions in the second monomer is thus higher, resulting in negative cooperativity.

This observation we make from the GNM is however slightly different from the NMR results that show activated motions throughout the whole protein upon the first binding. Later in this section we evaluate the average fluctuations of individual monomers within our model and show that this also suggests that the dynamics evolves in the way see in the iGNM model.

From the derived structure of the $\Delta\Delta G$ landscape we concluded that negative cooperativity can arise in a coupled dimer for a particular choice of parameters. We inferred that for negative cooperativity to arise the parameter α is most likely to be smaller than one. This corresponds to coupling between subdomains weakening upon the ligand binding. We also found from our exploration of the general model that $\Delta\Delta G$ is maximal when the loosening-tightening effect is present, suggesting that optimal design of a negatively cooperative system displays such a change in fluctuation amplitudes. Experiments have demonstrated that the loosening-tightening effect indeed occurs in CAP during the cooperative binding, strongly supporting our hypothesis [4]. In the following we want to use the remaining experimental results to determine if the dynamical structure of CAP is captured by our model, and if it is, to further localise CAP in the parameter space and gain more insight into the mechanism of its cooperativity.

We showed that the experimental value of $\Delta\Delta G = 4.7\,\mathrm{k_B T}$ can be recovered by including additional slow modes. In order to account for the favourable enthalpy change ($\Delta\Delta H = -1.8\,\mathrm{k_B T}$) we need to add fast modes as reviewed above. Enthalpy has been experimentally found in the CAP system to favor the second ligand binding, which corresponds to $\Delta\Delta H < 0$. By plotting Eq. 3.20 we can find the region of the parameter space with negative enthalpy.

The amplitude of the slow mode fluctuations are also identical to the non-enslaved case if $A = 0$ (Eq. 3.18). We localize the part of the parameter space with properties matching
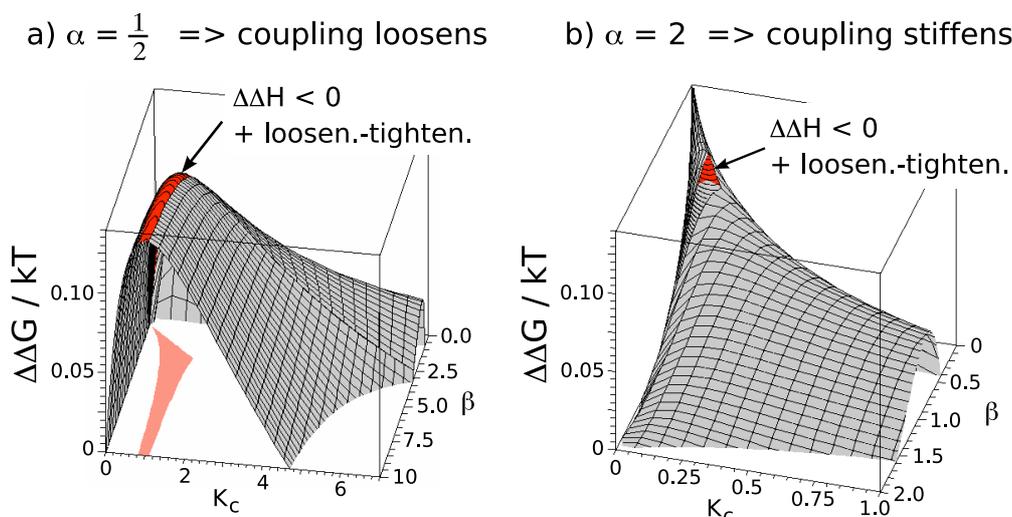
FIGURE 3.12: The allosteric free energy landscapes in the case of a single slow mode is coupled to a set of identical fast modes, for $\alpha = 1/2$, 2 with the area displaying the loosening-tightening effect plus $\Delta\Delta H < 0$ highlighted in red. The projection of the area into the K_c - β plane is shown in orange.

experimental results: $\Delta\Delta G > 0$, $\Delta\Delta H < 0$ and displaying the loosening-tightening effect upon binding. This parameter subspace also coincides with high allosteric free energy (see Fig. 3.12).

Fast fluctuations $\langle x_{f_i}^2 \rangle$ evaluated from our model cannot be compared directly to the experiment because the NMR experiments only measured fast motions of the backbone. Our model incorporates small structures such as side chains into $\langle x_{f_i}^2 \rangle$. A 40 ns molecular dynamics simulation performed by Li *et al.*, however reports on the fast motions of the whole molecule [141]. The rms deviation of the whole structure was found to decrease upon binding. This measure accounts for both side chain and backbone motions but is the best guideline available to us. We therefore add the decreasing fast fluctuations to the desired properties of our model. The fast mode fluctuations $\langle x_{f_i}^2 \rangle$ calculated with our model display a sequential tightening during the two binding steps for $\beta \gtrsim 1$ and sequential loosening for $\beta \lesssim 1$. Only in the case of $\alpha < 1$ does the area with $\Delta\Delta H < 0$ and the loosening-tightening effect stretch to large values of β (Fig. 3.12). This supports the hypothesis that the coupling between the CAP subunits is weakened upon the ligand binding.

As mentioned above, the results from iGNM indicate that the first cAMP binding quenches the motions in the interacting monomer but activates the motions in the

unbound monomer (Fig. 3.11). This contrasts with the NMR results of the group of Kalodimos showing that slow fluctuations increase evenly throughout the whole dimer [4].

We calculate mean fluctuations of the individual monomers $\langle x_i^2 \rangle$ from our coarse-grained model for comparison. The functions are relatively complex so in order to make the analysis more transparent we focus only on the case of $\alpha < 1$. Furthermore we assume that the ligand binding has local stiffening effect ($\beta > 1$), i.e. the internal monomeric spring constant becomes stiffer upon ligand binding. This effect is indicated by the elastic network models as well as by our own.

As expected if $\beta > 1$ then the mean fluctuations of the liganded monomer of the 1:1 state are smaller than of the unliganded monomer (1:1 refers to a singly liganded protein). The dimensionless ratio between the fluctuations of the individual monomers in the singly liganded state

$$q_{1:1} = \frac{\langle x_2^2 \rangle_{1:1}}{\langle x_1^2 \rangle_{1:1}}, \tag{3.21}$$

is plotted in Fig. 3.13. The ligand is bound to the monomer 1.

The amplitude of the fluctuations of the individual monomers increases when the first ligand binds for some values of parameters and decrease for others. The increase only occurs for $\alpha < 1$. In a small region of the parameter space motions in both monomers become larger. The dual increase may lead to the conclusion that we located the correct parameter subspace for CAP. However this region is characterised by small values of $\Delta\Delta G$ (Fig. 3.12) and the fluctuation amplitude of the doubly liganded protein is larger than any of the singly liganded protein. These two observations motivate further exploration of the parameter space.

Outside of this region the fluctuations of the liganded subunit of the 1:1 complex are reduced in amplitude compared to the apo state (as seen in iGNM). The amplitude of the motions of unliganded subunit are however increased in much larger subspace. In order to compare this subspace with the area of negative cooperativity we calculate the ratio of the average fluctuations of the unliganded monomer of the 1:1 state to the same monomer in the apo state, i.e.

$$q_{1:1,apo} = \frac{\langle x_2^2 \rangle_{1:1}}{\langle x_2^2 \rangle_{apo}}, \tag{3.22}$$

(a)

(b)
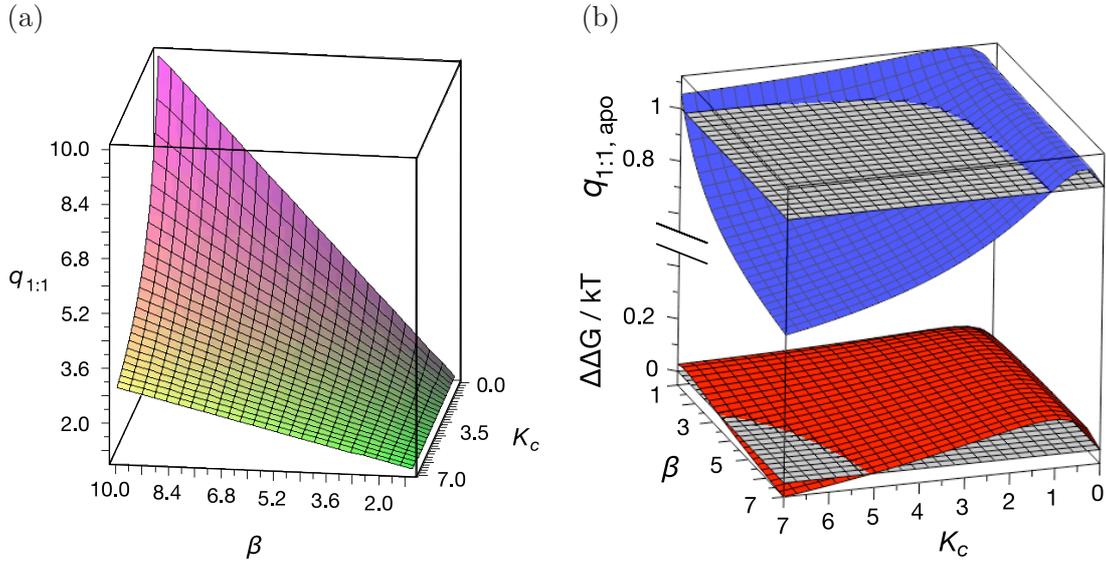
FIGURE 3.13: (a) The ratio $q_{1:1}$ of the average fluctuations in the unliganded and liganded monomer of the singly liganded protein. (b) Allosteric free energy landscape of a coupled dimer (in red) and the ratio $q_{1:1,apo}$ (Eq. (3.22)) plotted in blue. The region of positive $\Delta\Delta G$ is highlighted by the bottom grey plane and the region with increased fluctuations in the unliganded monomer of the 1:1 state is highlighted by the top grey plane. Both figures are plotted for $\alpha = 1/2$.

where the ligand is bound to the monomer 1. We observe that if $q_{1:1,apo} > 1$ then $\Delta\Delta G > 0$, in other words if the average fluctuations in the unliganded monomer of the 1:1 state increase compared to the apo state then negative cooperativity occurs (Fig. 3.13).

In the next step we compare the region of the parameter space with $q_{1:1,apo} > 1$ to the region of the parameter space identified previously as the location of CAP, i.e. the region with $\Delta\Delta G > 0$, $\Delta\Delta H < 0$ and the loosening tightening effect (red stripe in Fig. 3.12 a). Figure 3.14 a reveals that the CAP subspace completely falls into the area with $q_{1:1,apo} > 1$ (behaviour seen in iGNM). The fluctuations in the liganded subunit are also activated in a small region that is hidden in this view behind the red peak and is highlighted in a 2d plot in Fig. 3.14 b.

In conclusion, both iGNM and our model suggest that binding of the first ligand quenches the slow motions in the interacting subunit while activating them in the free subunit. Both models also indicate that the second binding reestablishes symmetry by suppressing the motions in the interacting subunit while negligibly increasing them in the bound monomer. The NMR data however show a rather uniform increase of the fluctuation

(a)

(b)
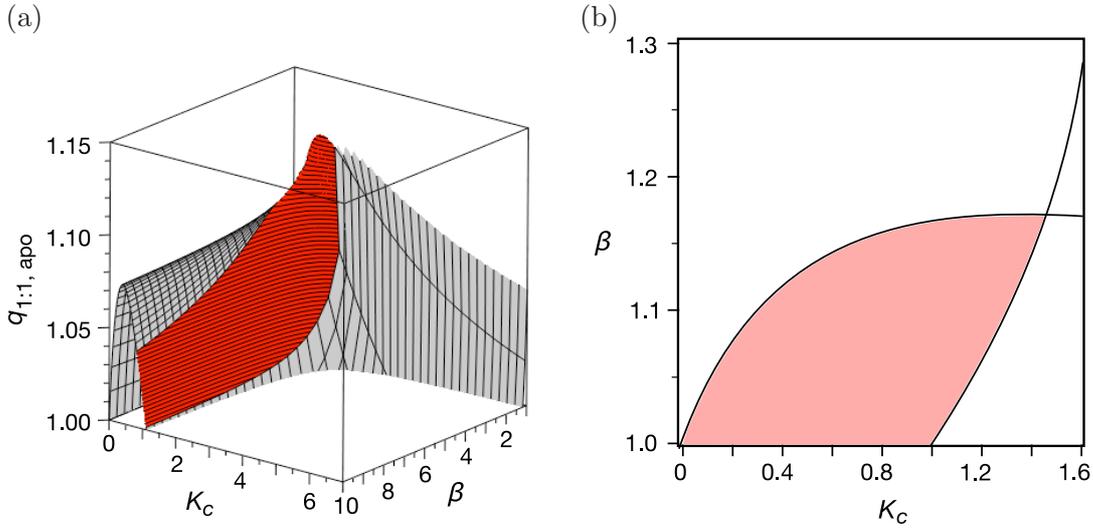
FIGURE 3.14: (a) The surface plot of the ratio $q_{1:1,apo}$ (Eq. (3.22)). The parameter subspace identified as a location of CAP is pictured in red, as in Fig. 3.12. This red region correlates with the increased amplitude of fluctuations in the unliganded monomer of the 1:1 state and suppressed fluctuations in the liganded monomer. (b) In a small region of parameter space, hidden behind the peak in the view in part (a), the average fluctuations in both subunits display an increase upon the first ligand binding. This region is highlighted in the 2d plot in pink. Both figures are plotted for $\alpha = 1/2$.

amplitude throughout the whole dimer (Fig. 3.8). We speculate that a critical review of the results and additional experiments may show that the dynamic pattern is modified by the ligand binding in the way the theoretical models suggest.

Let us return our focus to the prediction of the location of CAP in the parameter space regardless of the individual monomer fluctuations. We concluded that the parameter α is likely to be smaller than one and the remaining parameters defining the allosteric properties of CAP are probably of values within the small parameter subspace marked in red in Fig. 3.12 a. This is a very significant advance. However we found that the addition of fast modes alone does not capture the experimental magnitudes of $\Delta\Delta G$. This suggests that multiple slow modes are active in the allosteric effect alongside with the fast modes.

The most complex model we studied is composed of multiple slow and fast modes. The fast potential is assumed to depend on the superposition of the local slow modes, i.e. fast modes on the subunit 1 depend on the displacement $x_{s_1}^2 = x_{1,1}^2 + \ldots + x_{1,M}^2$. The allosteric free energy was evaluated, but the exact formula is not shown because of its length and complicated dependency on number of slow modes. Analysis of the result shows that

the characteristic properties of the one slow mode case are retained. Let us again denote $A = V_{f_0} k_v / k k_B T - k_k / 4k$ and set $A = 0$ for simplification. The free energy then equals the free energy of the slow modes only and is split into an enthalpic and entropic part. The enthalpic part is directly proportional to the number of enslaved fast modes and the region of parameter space with properties matched to experiment: $\Delta\Delta G > 0$, $\Delta\Delta H < 0$, the loosening-tightening effect and doubly suppressed fast modes spans the ridge of high $\Delta\Delta G$ in the subspace of $\alpha < 1$ (Fig. 3.12 a). The doubly suppressed fast modes now rule out completely the case of $\alpha > 1$.

Including more slow modes and amplifying the change in coupling constant induced by cAMP binding (choosing the value of α further away from 1) both increase the allosteric free energy. The experimental observations may be quantitatively recovered for the highly suggestive value of six global modes, $\alpha = 1/4$, $\beta = 8$ and $K_c = 1/2$. The number of fast modes can be fitted from the value of $\Delta\Delta H = -1.8 \, k_B T$. The form of the derived equation for $\Delta\Delta G$ is preserved from the one slow mode case (Eq. (3.18)), i.e. $\Delta\Delta H = N V_{f_0} K_v f(\alpha, \beta, K_c)$. We estimate the value of $V_{f_0} K_v$ to be $\lesssim k_B T$. For $V_{f_0} K_v = 0.1 \, k_B T$ and $N = 6$, $\alpha = 1/4$, $\beta = 8$ and $K_c = 1/2$ we need 15 fast modes per subunit to recover the $\Delta\Delta H$ value, for $V_{f_0} K_v = 0.5 \, k_B T$ and identical remaining parameters only three fast modes are required.

As noted in the case of the lac dimer [38], there are six mutual global modes of motion between two internally rigid domains (three relative translations, three relative rotations), suggesting that each monomer of CAP is composed of two semi-rigid subunits. This is confirmed by the Gaussian network model simulations where the β-sheet structure and the long central helix (C-helix) are observed to move as semi-rigid bodies. Their relative motion is also evident from the correlation maps. The β-sheets of each domain are highly correlated with each other and the small uncorrelated islands in the pattern correspond to the long β-hairpin that moves fast on its own. The mobility of the structure (Appendix A) increase with the increasing distance from the central helix, demonstrating that the structure moves with respect to the nearly stationary helices which are uncorrelated with the rest of the domain.

The intersubunit coupling is provided by the long C-helices. The role of the helices has been studied by Heyduk [139] and it was found that when the DNA binding domain (DBD) and the helix are missing, cAMP binding still occurs with the same affinity as in the full version of CAP, however the binding is non-cooperative. This corresponds to the case when coupling is not modified upon binding ($\alpha = 1$) and no cooperativity occurs.

When on the other hand the helix is present, but the DBD removed, the binding is tighter and more anti-cooperative than in the complete CAP. Popovych *et al.* showed that cAMP binding introduces a coil-to-helix transition in the untruncated version of CAP, where residues 125-136 are turned from a random coil in apo-CAP to alpha helix in the liganded CAP [142]. α-helices in the coiled-coil conformation interact more strongly than the random coils. This transition might therefore act against the coupling loosening present in the truncated version and might therefore reduce the strength of the cooperativity.

In conclusion, although current data still underdetermine the parameter set of the minimal quantitative coarse-grained model, it is still possible to conclude that somewhere in between a few and a few tens of fast modes are probably enslaved by the CAP dimer, a physically reasonable range

3.4   Oligomeric Proteins

In the first part of this chapter we focused on homodimeric proteins and evaluated the possible contributions of thermal fluctuations to the allosteric effect associated with symmetric binding. A large number of proteins are organized into higher homo-oligomeric forms, such as homotrimers, homotetramers etc. Homotropic allostery is observed in all types of oligomers, from tetramers such as haemoglobin [27] or glyceraldehyde phosphate dehydrogenase (GAPDH) [143] over mid-sized complexes like the 11mer TRAP [144] to proteins as large as 144mer haemoglobin from the leech Macrobdella decora [145]. Moreover, large proteins are notoriously difficult to study experimentally and it is thus highly probable that a large fraction of these phenomena has not been observed. A theoretical model that can suggest experimental tests may therefore be particularly important.

The model we proposed in the previous sections can readily be extended to model proteins with a larger number of protomers. The allosteric mechanism remains the same but the allosteric free energy can be evaluated for the second and every following binding step. This means that more types of overall allosteric effects arise. In principle, the system can be 1) increasingly cooperative or anti-cooperative, 2) the values of $\Delta\Delta G$ can fluctuate but retain the sign (all steps are positively or negatively cooperative) or 3) $\Delta\Delta G$ can change sign for some binding steps.

Another important difference is that the order in which ligands bind to the oligomer matters. This suggests an additional number of scientific questions. Can we predict the most probable binding pathway? Knowing the order of binding can we infer anything about the model parameters? How many independent binding pathways is there in an oligomer with a defined contact topology?

Most of these questions are not answered in this section but the general methodology for obtaining the allosteric free energies for each step is outlined. We first focus on the simplest oligomer with non-trivial binding order, the tetramer, and then suggest the generalization of this method to higher oligomers. We chose a tetramer because one of the most famous and most intriguing allosteric systems, the human haemoglobin A, is a tetramer. A recent study has shown that dynamics plays an important part in the allostery of haemoglobin [146] and here we aim to shed some light on the possible mechanism.

3.4.1 Model of a Homotetramer

Let us consider a protein composed of four identical subunits arranged in a circular manner as shown in Fig. 3.15 a, using the example of haemoglobin. Four identical ligands bind to the protein as one ligand per protomer. In this section we only discuss the analogy of the simplest model of the homodimer introduced above, i.e. one internal mode is assigned to each subunit of the tetramer and each subunit is coupled to its two nearest neighbors. The internal mode is defined by a spring constant k and the elastic coupling by a constant k_c. The model is sketched in Fig. 3.15.

Similarly to the previous section the change in the mass of the protein upon ligand binding is considered negligible and the mass part of the Hamiltonian is neglected. The coarse-grained model of the unliganded state is described by the Hamitonian

$$\mathcal{H} = \mathbf{x^T}\mathbb{K}_0\mathbf{x}, \tag{3.23}$$
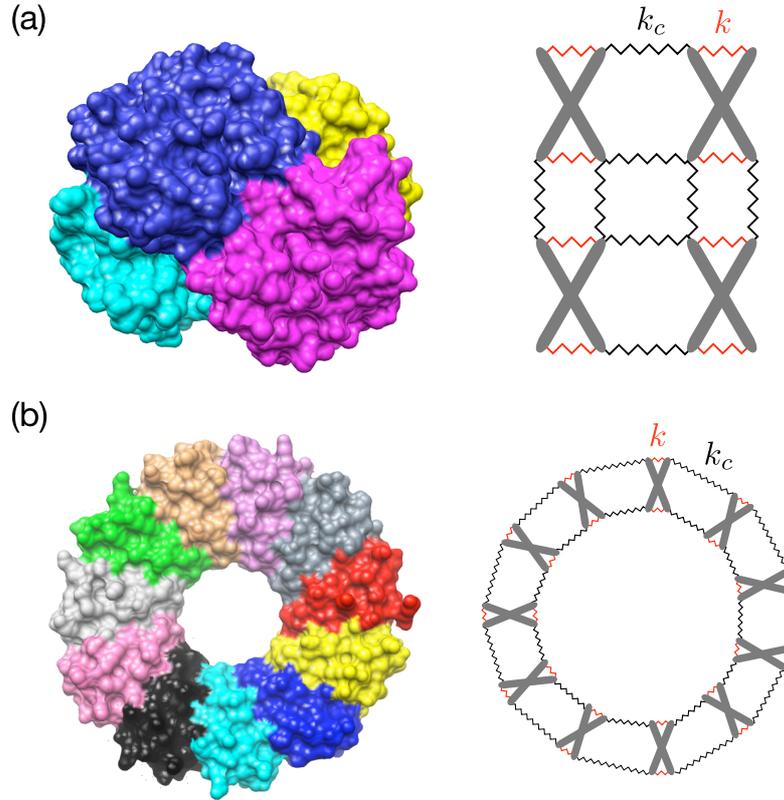
FIGURE 3.15: The surface model of (a) haemoglobin (pdb entry 1A3N) and (b) ring protein TRAP (pdb entry 1WAP) and their respective coarse-grained models. Each subunit is assigned a single internal mode, shown here schematically as a scissor-like movement of the grey rods and each subunit is coupled to its two nearest neighbors. The internal motions of each subunit are defined by an identical spring constant k and the coupling by a constant k_c.

where $\mathbf{x}$ is a vector of the coordinates of the individual subunits and $\mathbb{K}_0$ is a square 4×4 matrix

$$\mathbb{K}_0 = \begin{pmatrix} k + 2k_c & -k_c & 0 & -k_c \\ -k_c & k + 2k_c & -k_c & 0 \\ 0 & -k_c & k + 2k_c & -k_c \\ -k_c & 0 & -k_c & k + 2k_c \end{pmatrix} \tag{3.24}$$

The ligand binding is assumed, as in the previous section, to alter the internal breathing mode of the interacting subunit by factor β and the coupling to the nearest neighbours
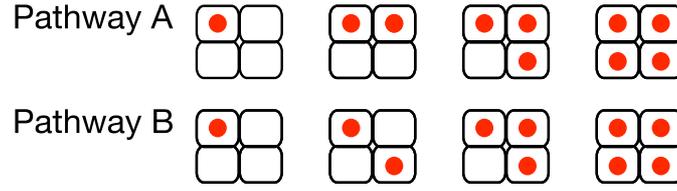
FIGURE 3.16: Two distinct binding pathways of four substrate molecules (in red) to a tetramer.

by factor α. The matrix of the singly liganded protein is then

$$\mathbb{K}_1 = \begin{pmatrix} \beta k + 2\alpha k_c & -\alpha k_c & 0 & -\alpha k_c \\ -\alpha k_c & k + (\alpha+1)k_c & -k_c & 0 \\ 0 & -k_c & k + 2k_c & -k_c \\ -\alpha k_c & 0 & -k_c & k + (\alpha+1)k_c \end{pmatrix} \tag{3.25}$$

The tetramer can fill its four binding sites in two distinct pathways sketched in Fig. 3.16. The difference is in the location of the second binding. This can occur either next to the occupied binding site (we refer to this as a pathway A) or on the diagonal (pathway B).

The elastic matrices $\mathbb{K}_{2A}$, $\mathbb{K}_{2B}$, $\mathbb{K}_3$ and $\mathbb{K}_4$, of the protein with two and more ligands bound are found by modifying the appropriate spring constants in the matrix Eq. (3.24). Three allosteric free energies are evaluated for each pathway. The notation for the energies is chosen as follows

$$\Delta\Delta G_{2A} = (G_{2A} - G_1) - (G_1 - G_0) \tag{3.26}$$

where G_{2A} refers to the free energy of doubly liganded protein on the pathway A.

As for the example of a homodimer, we divide the parameter space $\{\alpha, \beta, K_c\}$ into two subspaces, subspace of $\alpha < 1$ in which the interaction between protomers becomes weaker upon the ligand binding and subspace $\alpha > 1$. The free energy landscapes for the two pathways in the subspace $\alpha < 1$ and $\alpha > 1$ are shown in Fig. 3.17 and 3.18 respectively.

The subspace $\alpha < 1$ is characterised by alternating positive and negative cooperativity. On the pathway A, the second and fourth ligand binding are increasingly anti-cooperative with α diverging from 1. The third ligand binding is weakly cooperative. On the path B, the second ligand binds with increased affinity compared to the first, but

is then followed by a significantly anti-cooperative third binding step and only negligibly cooperative fourth. Positive cooperativity for all binding steps can actually be obtained in this subspace, but only for α approaching unity. The values of free energy differences in this case are however very small, in particular for the last two ligand associations ($\Delta G_{3,4} < 0.1\,k_B T$).

Similarly to the homodimer, positive cooperativity prevails in the subspace $\alpha > 1$, although individual binding steps are occasionally weakly anti-cooperative. Path A displays mainly sequential cooperativity. Only for values of $\beta \sim 1$ the free energies associated with individual binding steps, ΔG_i, become very small and some binding steps weakly anti-cooperative. Path B is characterised by positive cooperativity occurring upon the second and last ligand binding. The third ligand binds mostly with reduced affinity and the negative cooperativity becomes amplified with increasing value of α.

The model can be used to calculate the monomeric fluctuations as

$$\langle x_i^2 \rangle = \frac{1}{Z} \iint \mathrm{d}x_1 \ldots \mathrm{d}x_4\, x_i^2 \exp\left(-\frac{\mathcal{H}(x_1,\ldots,x_4)}{k_B T}\right), \tag{3.27}$$

where $\mathcal{H}$ and Z are the Hamiltonian and the partition function of the appropriate binding state. We obtain fluctuations of four protomers for each state as a function of α, β and K_c. If $\alpha > 1$ and $\beta > 1$ the amplitude of the protomers' fluctuations decreases upon the sequential ligand binding on both paths. For $\alpha < 1$ and $\beta < 1$ tightening loosening effects occur at some binding stages. Relative fluctuations of neighboring protomers can be found similarly.

The model introduced here can be extended to include fast and multiple slow modes on each protomer. This is achieved with the methods introduced in Sec. 3.2. The additional modes can be included if we wish to fit the model with the data from a particular system for which the contribution of dynamics has been measured. Because we do not have such data available we focus on more qualitative analysis of the simple model. We find that this type of analysis nevertheless results in a valuable contribution to understanding of allosteric effects in higher oligomers. We illustrate this on an example tetramer, the haemoglobin.
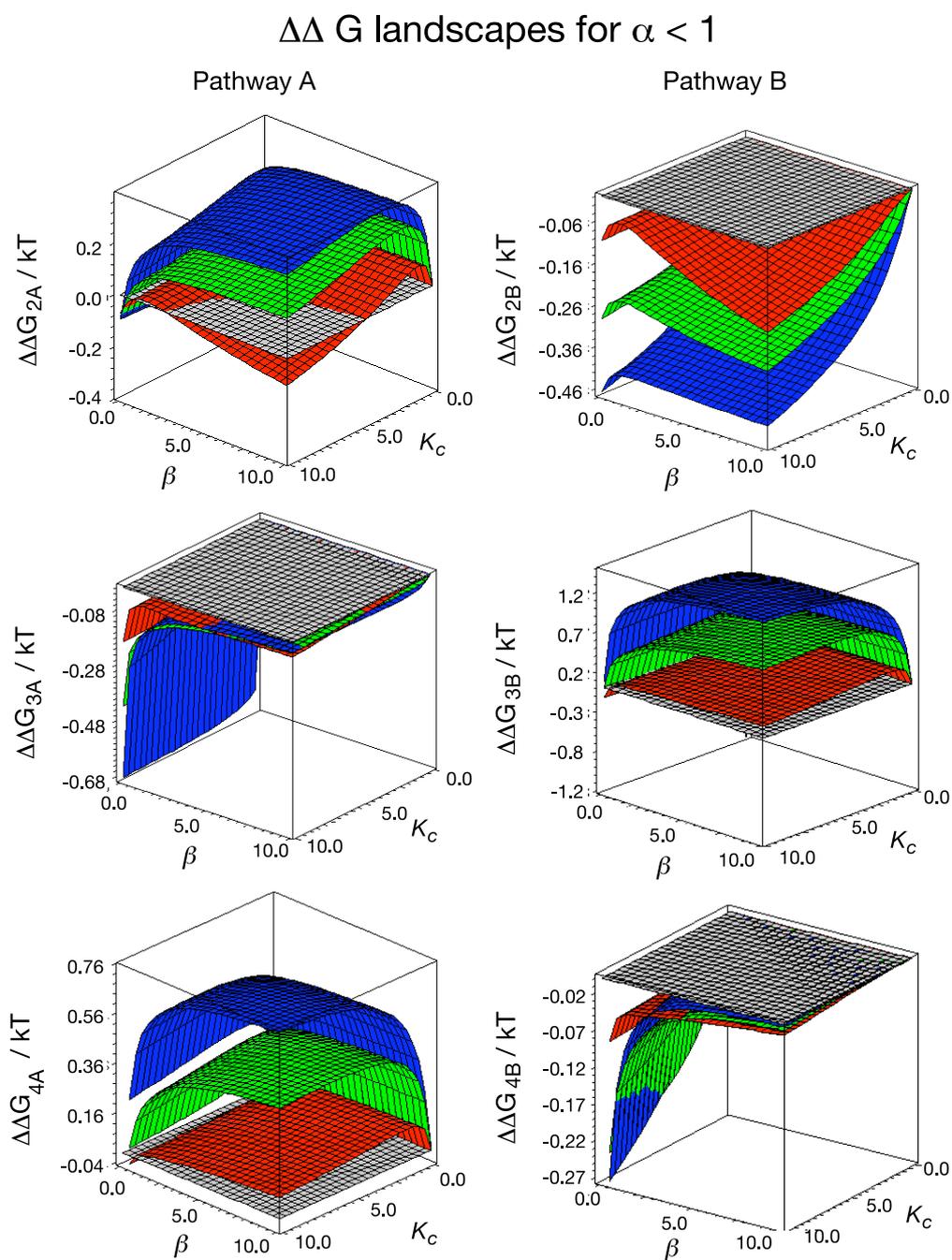
FIGURE 3.17: Allosteric free energy associated with the second, third and fourth ligand binding for the values of $\alpha < 1$. The left column corresponds to the path A and left to path B, the colors reflect the value of α. The free energy landscape for $\alpha = 1/2$ is shown in red, for $\alpha = 1/5$ in green and for $\alpha = 1/10$ in blue.
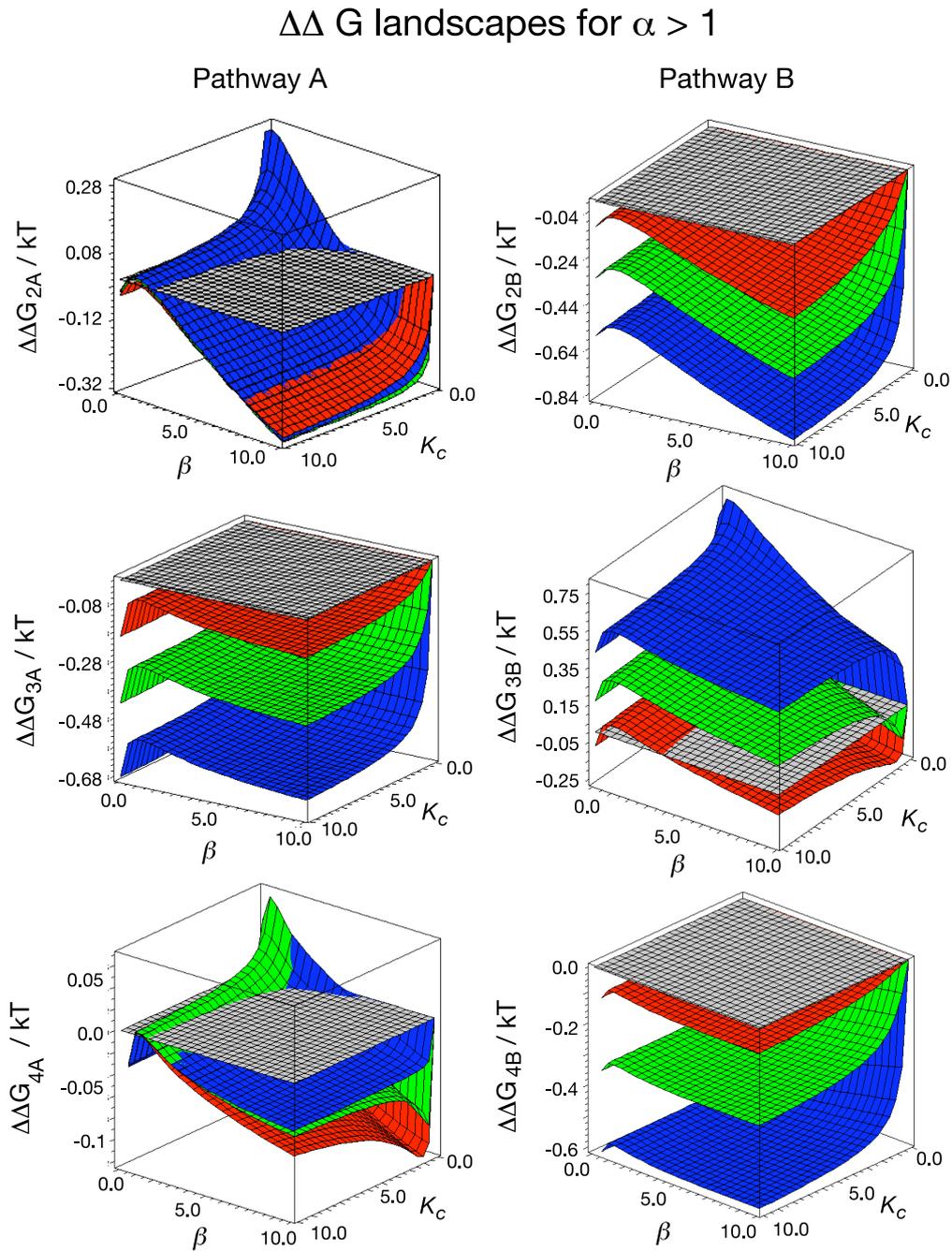
FIGURE 3.18: Allosteric free energy associated with the second, third and fourth ligand binding for the values of $\alpha > 1$. The left column corresponds to the path A and left to path B, the colors reflect the value of α. The free energy landscape for $\alpha = 2$ is shown in red, for $\alpha = 5$ in green and for $\alpha = 10$ in blue.
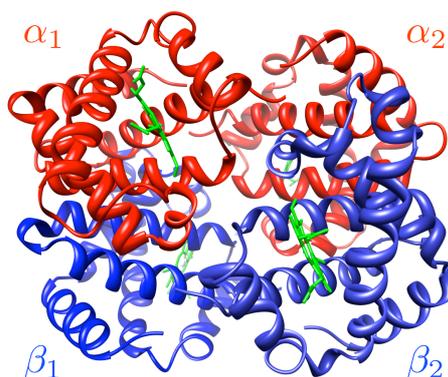
FIGURE 3.19: The ribbon diagram of the structure of human haemoglobin A (pdb entry 1A3N). The α subunits are displayed in red, β subunits in blue and the haeme groups in green sticks.

3.4.2 Example Tetramer: Haemoglobin

Mammalian haemoglobin (Hb) is arguably the most famous and most studied protein and in particular its allosteric properties have been attracting attention ever since its discovery in 1904. It is composed of four subunits, two α and two β arranged with an approximate tetrahedral symmetry (Fig. 3.19). The α and β subunits are structurally and evolutionarily closely related and can thus be considered identical for coarse-graining purposes. Each of the subunits contains a haeme group, which binds a molecule of oxygen. The four oxygen molecules bind in a cooperative manner with the Hill coefficient of 2.5 [147] (for the meaning of the Hill coefficient see Sec. 1.2). Other ligands can also associate with haemoglobin and influence its affinity for oxygen. Here we focus on the homotropic allostery that occurs during the oxygen binding alone, details about the haemoglobin's heterotropic allostery can be found e.g. [147].

The homotropic allostery is widely believed to proceed through structural changes discovered originally by Max Perutz and later refined with a number of experimental techniques [27]. The $\alpha\beta$ dimers of oxygen saturated haemoglobin (oxy-haemoglobin) are rotated by ~ 15° compared to the oxygen free form (deoxy-haemoglobin). A number of models have been published explaining the allosteric mechanism, some of which are introduced in Chap. 2, others have been reviewed in [27]. Recently however Laberge and Yonetani critically reviewed all existing models and found that they are all based on problematic and outdated assumptions [146]. They used logical arguments and results of their earlier molecular dynamics study [148] to suggest that structural changes are

not sufficient to explain the homotropic allostery observed in haemoglobin and pointed out that hitherto neglected dynamics may play a significant role.

The study that Laberge and Yonetani performed consists of 6 ns MD simulation of two binding states of haemoglobin A: the oxygen free (deoxy-Hb) and the oxygen saturated form (oxy-Hb). They have found that the oxy-Hb has significantly reduced fluctuation amplitude compared to the deoxy-Hb [148]. The importance of fluctuations was further highlighted by the simulations in the presence of a heterotropic allosteric effector 2,3-Diphophoglycerate (DPG). DPG binds non-covalently to the central cavity of Hb, a site distant from the haemes, and substantially reduces the cooperativity of oxygen binding. X-ray crystallography reveals no major change in the average conformation of Hb upon DPG binding [149]. The prevailing opinion is that DPG freezes the haemoglobin in its deoxy conformation and so prevents the subsequent oxygen binding [146].

The MD simulations confirm the negligible change in the structure, but observe a substantial change in the DPG bound Hb dynamics. Whereas the deoxy-Hb-DPG fluctuates, as expected, with slightly smaller amplitude than DPG free form, the analysis of the oxy-Hb-DPG gives a surprising result. Namely, DPG bound oxy-Hb has larger fluctuation amplitude than other binding states. Laberge and Yonetani conclude that the increased flexibility acts against the second and following oxygen bindings, i.e. against cooperativity. Furthermore the simulations show that only the 10 lowest frequency modes are perturbed by the DPG binding [148].

A number of experimental studies have been performed with the aim to determine precise binding free energies of each oxygen binding. The strong cooperativity prevents characterization of the intermediate binding states directly and the constants thus have to be found from fitting the binding curve into a particular binding model. Most commonly the Adair's equation (1.22) is chosen as a model. Most recently Yonetani *et al.* [147] determined the Adair's apparent binding constants at pH 7

$$K_1 = 7.2 \cdot 10^{-3} \qquad K_2 = 1.05 \cdot 10^{-2}$$
$$K_3 = 1.15 \cdot 10^{-2} \qquad K_4 = 2.33 \cdot 10^{-2}. \qquad (3.28)$$

In order to calculate allosteric free energies we have to convert these apparent or stepwise equilibrium constants into the equilibrium constants of the individual bindings (see

Sec. 1.1.1). Using the equations (1.10) and (1.12) we obtain

$$\Delta\Delta G_2 = -1.4\,k_B T$$
$$\Delta\Delta G_3 = -0.9\,k_B T$$
$$\Delta\Delta G_4 = -1.7\,k_B T. \tag{3.29}$$

Ackers have used a different method for the free energy evaluation with the additional aim to characterise all possible binding microstates [150]. A microstate in this context refers to haemoglobin with a certain number of oxygen molecules bound to specified subunits. The method employed was based on thermodynamic linkage, which implies that the difference in the free energy of dissociation of two tetrameric microstates into the corresponding $\alpha\beta$ dimers is equal to the difference in free energy upon binding of the ligands to the tetramers and the dimers. Because binding of oxygen to the dimers is non-cooperative and has nearly the same affinity for oxygen as the free tetramer, the free energies obtained are equal to the free energies associated with ligating each microstate. Using this method they obtained free energies of 10 distinct microstates, the results are shown in Table 3.1.

Ligation microstate	$\Delta G/k_B T$	Ligation microstate	$\Delta G/k_B T$
$\alpha_1\alpha_2\beta_1\beta_2$	0	$\alpha_1 O_2\alpha_2 O_2\beta_1\beta_2$	11.3 ± 0.3
$\alpha_1 O_2\alpha_2\beta_1\beta_2$	4.7 ± 0.5	$\alpha_1\alpha_2\beta_1 O_2\beta_2 O_2$	11 ± 0.3
$\alpha_1\alpha_2\beta_1 O_2\beta_2$	4.7 ± 0.5	$\alpha_1 O_2\alpha_2\beta_1 O_2\beta_2 O_2$	11.5 ± 0.3
$\alpha_1 O_2\alpha_2\beta_1 O_2\beta_2$	8.3 ± 1.3	$\alpha_1 O_2\alpha_2 O_2\beta_1 O_2\beta_2$	11.5 ± 0.3
$\alpha_1 O_2\alpha_2\beta_1\beta_2 O_2$	11.2 ± 0.7	$\alpha_1 O_2\alpha_2 O_2\beta_1 O_2\beta_2 O_2$	10.5 ± 0.2

TABLE 3.1: Free energy associated with ligating 10 distinct microstates of haemoglobin. Adapted from [150].

We observe that only in the route through the state $\alpha_1 O_2\alpha_2\beta_1 O_2\beta_2$ all binding steps are positively cooperative. The allosteric free energies on this path $\Delta\Delta G_2 = -1.1\,k_B T$, $\Delta\Delta G_3 = -0.4\,k_B T$ and $\Delta\Delta G_4 = -4.2\,k_B T$ are qualitatively in agreement with those found by [147] and shown above (Eq. (3.29)).

The lower value of free energy associated with ligating the microstate $\alpha_1 O_2\alpha_2\beta_1 O_2\beta_2$ compared to the other doubly liganded tetramers is inconsistent with the classical MWC model, commonly used to explain cooperativity of haemoglobin [27]. When compared with Fig. 3.19 we notice that binding through this state corresponds to what we denoted as path A, the only fully cooperative path in our model. The other microstates that

correspond to the path A ($\alpha_1 O_2 \alpha_2 O_2 \beta_1 \beta_2$ and $\alpha_1 \alpha_2 \beta_1 O_2 \beta_2 O_2$) however do not display lower free energy, suggesting that the reason for lowered free energy penalty could also be the larger number of intra-dimer than inter-dimer contacts.

Our model demonstrates that positive cooperativity can arise in a coupled tetramer purely from modified fluctuations. At this stage the model is not detailed enough and the experimental and computational data on the intermediate states (with 1 - 3 ligands bound) are too limited to allow an equivalent determination of elastic parameters, as was achieved for CAP. Our data however strongly suggest that haemoglobin is located in the subspace $\alpha > 1$, $\beta > 1$ and binds oxygen in the order of pathway A (Fig. 3.16). The path A is characterized by sequential cooperativity for $\alpha > 1$ and $\beta > 1$. Furthermore, in agreement with Laberge and Yonetani [148], the monomeric fluctuations are sequentially reduced upon each binding.

Additionally both experimental results suggest that the last step is associated with the largest cooperative free energy. In our model the last step is usually significantly less cooperative than the previous steps. We can speculate that since the oxygen binding is accompanied by a structural change (relative rotation of the $\alpha\beta$ dimers by ~ 15°), this change occurs upon the last binding and results in enthalpically favourable interaction reflected in the values of free energy.

The partially liganded intermediate states of Hb have been studied before with crystallographic [151, 152] and spectroscopic [153] methods. The results suggest that the partially liganded states are characterised by intermediate structural changes, such as they retain their T-state quaternary structure but the liganded subunits have modified R-like tertiary structures. It is thus plausible that the last ligand binding is accompanied by particularly enthalpically favourable structural rearrangements.

Haemoglobin however is not an ideal model system because the inter-dimer contacts are substantially different to those at the dimer interface. In the future the model could either be modified to account for different strength of interaction between the subunits or the predictions could be tested on a different model system. One possible candidate was mentioned in the introduction to this section, the glyceraldehyde phosphate dehydrogenase (GAPDH). More details about this system are given as a part of future work in Chap. 6.

3.4.3 Model of a Circular Oligomer

The models of a dimer and a tetramer can be generalised to include N subunits. Here we outline the general features of such model. Let us consider analogous arrangement as was used on the example of tetramer. A circular protein consists of N identical subunits each of which is assigned one internal breathing mode and is coupled to its two nearest neighbours (Fig 3.15). Again, the internal mode is defined by a spring constant k and the elastic coupling by a constant k_c.

The model is mathematically defined by a Hamiltonian, which can be written in a matrix form involving a square $N \times N$ matrix. Similarly to the previous section the change in the mass of the protein upon ligand binding is considered negligible and the mass part of the Hamiltonian is therefore neglected. The unliganded state is defined by

$$\mathcal{H} = \mathbf{x^T} \mathbb{K}_0 \mathbf{x}, \tag{3.30}$$

where $\mathbf{x}$ is a vector of the coordinates of the individual subunits and $\mathbb{K}_0$ is a square $N \times N$ matrix

$$\mathbb{K}_0 = \begin{pmatrix} k + 2k_c & -k_c & 0 & \cdots & 0 & -k_c \\ -k_c & k + 2k_c & -k_c & 0 & \cdots & 0 \\ 0 & -k_c & k + 2k_c & -k_c & & \vdots \\ \vdots & 0 & -k_c & \ddots & \ddots & 0 \\ 0 & \vdots & & \ddots & \ddots & -k_c \\ -k_c & 0 & \cdots & 0 & -k_c & k + 2k_c \end{pmatrix}. \tag{3.31}$$

The matrix is tridiagonal with the exception of the non-zero elements in the top right and bottom left corner coupling the first and last residue in the circle.

The assumption of two nearest neighbours is more appropriate for some proteins than for others. For example each subunit of the ring protein TRAP is clearly in contact with only two other subunits. Other proteins, such as large haemoglobins are organized in complicated nested structures and this assumption may need to be reevaluated from case to case. The different contact topology is implemented by changing the structure of the tridiagonal matrix Eq. (3.31) by converting the zeros corresponding to the contacts into the appropriate coupling constants and vice versa.

The ligand binding can be assumed, as in previous sections, to alter the internal breathing mode of the interacting subunit by factor β and the coupling to the nearest neighbours by factor α. The matrix is then modified accordingly. For $N > 3$ the order of ligand bindings becomes important. We have analysed the two distinct pathways that arise for a tetramer, the number of pathways however grows quickly with the increasing N; there are four distinct pathways for a pentamer and 27 for a sextamer.

The number of distinct ligand arrangements for a given number of ligands can be obtained for a general oligomer. The number depends on the parity of the number of the oligomer's binding sites N. Let us sketch the calculation for the first few binding events. Independently of the parity, there is only 1 binding state with one ligand bound, two ligands can bind in $(N-1)/2$ indistinct ways if N is odd and in $N/2$ ways if N is even.

The third binding is already more complicated and results in

$$\sum_{i=0}^{\left[\frac{N-2}{6}\right]} \frac{N-1-6i}{2} + \sum_{j=0}^{\left[\frac{N-5}{6}\right]} \frac{N-5-6j}{2} \tag{3.32}$$

arrangements if N is odd. The square brackets in the sum's upper limit denote the nearest smaller integer, i.e. if $(N-2)/6 = 1.8$ then $[(N-2)/6] = 1$.

Similarly if N is even then three ligands can bind in

$$\sum_{i=0}^{\left[\frac{N-2}{6}\right]} \frac{N-1-6i}{2} + \sum_{j=0}^{\left[\frac{N-4}{6}\right]} \frac{N-4-6j}{2} \tag{3.33}$$

distinct ways. The total number of pathways is obtained as a product of all possible ligand arrangements for all sequential binding states.

The general expression becomes quickly complicated and the total number of distinct paths large, e.g. there are $5^2 \cdot 15^2 \cdot 18^2 \cdot 25^2 = 1.14 \cdot 10^9$ paths for a circular 11-mer.

The allosteric free energy can be evaluated for the second and then every following binding step in the same manner as for the tetramer. Again, the system can be increasingly allosteric (the absolute value of $\Delta\Delta G$ increases with every step), the values can fluctuate but retain the sign (all steps are positively or negatively cooperative) or can change sign for some binding steps. This depends on the chosen values of the parameters (α, β and K_c).

3.5 Summary and Conclusions

We have attempted to add to the understanding of the allostery of coupled dimers by constructing a simple but intuitive coarse-grained model based on the basic thermo-dynamic principles of ligand binding and protein dynamics. We derived a model that describes the propagation of the allosteric signal in a coupled dimer purely via slow global motions. We have shown that such a model can account for positive and negative allostery. This is achieved by fine-tuning of three parameters.

The parameter space is naturally divided into two subspaces ($\alpha > 1$ and $\alpha < 1$), each of which supports a different type of cooperativity. In the subspace where coupling becomes stronger upon binding ($\alpha > 1$) the system is very likely to be positively cooperative. In the opposite case, when the coupling becomes weaker upon binding ($\alpha < 1$), the remaining parameters would need to reach unphysical values if a significant positive cooperativity was to occur.

The relative fluctuations were evaluated for the homodimer and four distinct types of responses to the consecutive binding were found. When mapped onto the allosteric free energy landscape, predictions can be made as to what type of response is likely to occur for different types of cooperativity. An anti-cooperative system is expected to display a loosening-tightening effect whereas the fluctuation amplitudes of a positively cooperative protein are most likely to be suppressed by each binding.

The model containing one slow mode only is very instructive, however the magnitudes of the allosteric free energy are significantly smaller than in real systems. We therefore speculated that more slow modes are active in the allosteric signaling. We extended our model to include the extra modes and indeed found that values of $\Delta\Delta G$ are noticeably amplified. Values of several $k_B T$ corresponding to common biological systems were recovered for 5-10 slow modes. The character of the relative fluctuations was preserved from the single mode model.

We then validate our approach on a test case homodimer: the catabolite activator protein (CAP). We focused on explaining the internal mechanism and the origin of the thermodynamic parameters measured in experiments. From the findings of the first section we knew that slow, global modes are responsible for the free energy value but on their own produce a purely entropic effect. The value of $\Delta\Delta G$ increases with the number of slow modes involved. In order to account for the compensating enthalpic

and entropic parts fast modes were added to the system using the method of Hawkins and McLeish [42]. Fast modes, despite being localized, can contribute to the allosteric signaling as "enslaved" by the slow modes. They are responsible for splitting the free energy coming from change in dynamics into an enthalpic and entropic part. The extent of this split is determined by the number of enslaved fast modes, the more fast modes the larger the compensating enthalpic and entropic terms.

According to experiments and simulations the overall change in enthalpy $\Delta\Delta H < 0$ and fast fluctuations decrease during the two binding steps. These results, along with the loosening-tightening effect displayed by the slow modes, were captured by the full model containing multiple slow and fast modes. It also enabled us to localize CAP in the model's parameter space. The examination of the allosteric free energy landscapes suggests that $\alpha < 1$, i.e. the coupling between subunits becomes softer upon cAMP binding. The other requirements overlap in a small region of the parameter space highlighted in red in Fig. 3.12 a. This region covers a narrow strip of the free energy landscape with the highest values of $\Delta\Delta G$, a feature that is preserved also when multiple slow modes are introduced. Furthermore we recover observed calorimetric values quantitatively in the case of six global and ~10-20 fast modes per subunit. The case of six internal modes is very suggestive because there are six mutual global modes of motion between two internally rigid domains (three relative translations, three relative rotations). The CAP subunits do indeed contain two principal units (the long α-helix and the β-sheet structure) as demonstrated by performing a GNM simulation. It should not prove excessively difficult to identify these structures experimentally even though NMR measurements are currently mapped onto spatial, rather than modal, dynamics. In addition, the change in fluctuations seems to be optimized for the maximum anti-cooperative effect.

We elucidated the effect, puzzling at first sight, where binding of two identical ligands to a completely symmetric dimer has entirely different consequences. We have also shown that a change in protein dynamics can produce a non-zero enthalpy change and suggested how measured thermodynamic parameters can be interpreted. They indicate how many slow and fast modes are being harnessed for the allostery and how the local stiffnesses change. The importance of the coupling between the subunits of a dimer has been highlighted and the different extent of the cooperativity in truncated and complete version of CAP has been explained.

Employing our model as an analytical tool of current experimental data allows us to make new predictions and to suggest new experiments. Specifically we expect to find

that coupling between subunits weaken on cAMP binding, that two structures dominate the global dynamics and about 10-20 local structures couple to the global fluctuations. However the exact determination of the parameters relies on either new analysis of available data or new experimental and/or computer simulation results.

In the second part of this chapter we outlined an extension of the homodimer model to a general oligomer. We focused on the example of a tetramer due to the potential application to the arguably most famous allosteric protein, the haemoglobin. The model of the tetramer was comparatively simpler than that of a dimer because it included only one internal breathing mode per monomer and omitted the fast modes. This simple model nevertheless demonstrates that complex allosteric behavior can arise in a coupled tetramer, with the sign and degree of cooperativity varying between the binding steps depending on the values of elastic parameters.

Many patterns found in the example of homodimer were preserved in the model of a tetramer. The homo-tetramer tends to display positive cooperativity if $\alpha > 1$ and is largely anti-cooperative if $\alpha < 1$. Loosening tightening events are also observed in some binding stages but they are almost exclusively limited to the subspace of $\alpha < 1$.

The fit of elastic parameters achieved for CAP however could not be repeated due to the lack of details in our model and absence of experimental characterization of the intermediate binding states of haemoglobin (1-3 ligands bound to Hb). The model however strongly suggests that haemoglobin chooses to bind a second oxygen to the nearest subunit of the occupied monomer, an observation that agrees with experimental results [150]. The experimental data however display a preference to bind to the nearest neighbor on the same dimer, a resolution not implemented in our model.

Partially liganded Hb appears to have structural features of both, the oxy and the deoxy form. This finding is in disagreement with the classical MWC model and suggests that more than one mechanism is taking place during the cooperative binding. We propose that during the first three binding steps the dynamic changes could be responsible for significant amount of allosteric free energy and the fourth binding could give rise to enthalpically favourable interactions reflected in the large cooperative free energy.

We finished the chapter by outlining the evaluation of dynamic allostery in a general oligomer with subunits arranged in a ring. We found that in this case the number of distinct paths in which ligands can bind to the protein increases exponentially with

the number of subunits. This complicates the analysis and motivates different types of scientific questions.

Chapter 4

Allostery in the Elastic Rod Model of DNA

Ever since its discovery, DNA has been attracting attention not only due to its importance as a storage of genetic information but also thanks to its apparent simplicity. Its repeating structural units are assembled in such a fashion that on medium length scales (tens of base pairs or nanometers) DNA looks and behaves like an elastic rod, an unusually simple structure for a biomolecule [154].

In this and the following chapter we address and critique rod-like models of DNA in terms of their ability to account for propagation of allosteric signals across these length scales. Despite the large amount of experimental data and relatively established and tested theoretical description of its mechanical properties the role of these properties in allostery has not been subjected to many studies. We use theoretical tools of coarse-grained molecular biophysics and experimental results on DNA itself to evaluate the possible allosteric effects resulting from change in equilibrium fluctuations. We motivate and parameterise our model by a benchmark example of dynamic allostery in DNA: the DNA/Hoechst 33258 system [101, 155].

DNA offers a unique opportunity for studying allostery in a medium with uniform and well-understood elastic properties. The thermal motion of DNA is composed of a large number of spatially delocalised modes that all have to be taken into account in a realistic model of dynamic allostery. This contrasts with a typical protein in which the separation of regions of high and low elastic modulus leads to a separation of modes in the frequency

spectrum. The dominance of only few low frequency, global modes is then reflected in the coarse-grained models employed in the study of their dynamic allostery [38, 42, 126]. The stiff regions are coarse-grained into rigid blocks and the blocks connected by harmonic springs. The spring rigidities are altered upon the ligand binding which is equivalent to modifying selected frequencies of the vibrational spectrum, but keeping the structure of the modes in the spectrum either completely unchanged [41, 42] or only slightly modified [38, 126].

Hawkins and McLeish studied dynamic allostery in alpha helical coiled-coils modeled as two interacting elastic rods [41]. This can be seen as a similar system with uniform elastic properties. The authors however focused on one dominant mode affected by the binding, the sliding of the rods. The binding events were then simulated as point-like "clamping" of the rods thus influencing only one effective spring constant and resulting in unmodified structure of the normal modes.

Although the main motivation for the present study is the DNA molecule, we note that biofilaments that can be modeled as a thermally fluctuating elastic rods are ubiquitous in cells. Structures that have been previously coarse-grained in this way include RNA, cytoskeletal fibers such as actin or microtubules and long coiled-coil structures such as the dynein stalk. These cellular components are constantly interacting with their environment and allosteric effects are likely to play a part. Our study may therefore have potential implications for these systems.

4.1   Dynamic Allostery in DNA

DNA performs its biological functions via interaction with proteins. Proteins are responsible for packing DNA into chromosomes and they promote and assist transcription and replication. They rarely act individually, each binding event usually influences a whole range of other reactions, some of which may occur far away from the original event. Typically, binding of a protein can affect a gene transcription on a different segment of DNA [156]. The influence is generally exerted via two mechanisms: firstly, the proteins can interact directly, e.g. lac repressor binds to the lac operon region of DNA and sterically prevents the RNA polymerase from binding to DNA and transcribing the gene [125]. Secondly, the DNA molecule can become involved in transmitting the signal

between two proteins that do not directly interact [157–159], or where the interaction is weak [160, 161].

We are interested in the second group of cooperative interactions where DNA participates in the propagation of the signal. In the majority of known cases the DNA is observed to undergo a conformational change during the signal transmission; examples include form transition [157, 162], introduction of over- or undertwist [158, 163], sharp bending of the DNA [159], or other modifications [160, 161]. The structural change is often considered the sole mechanism of the signal propagation. However, dynamics also play an important role in the allosteric effects (see Chap. 1). The DNA molecule is in a thermal environment and any change of binding or structural state will inevitably disrupt its vibrational spectrum. The disruption is carried along the molecule and contributes to the allosteric effect. Here we investigate the role of bending and twisting modes of DNA motion and find if and to what degree they can contribute to allostery.

The role of thermal fluctuations in multiple binding to DNA has been theoretically investigated before [164–166]. Rudnick and Bruinsma investigated cooperative binding of a pair of proteins that introduce a significant bend to DNA [164]. They calculated enthalpic and entropic contributions to the free energy associated with the interaction of the two proteins mediated by the DNA and found that the contributions are non-zero only in the presence of tension in the DNA. The effective potential arising from the enthalpic contributions to the free energy can be either attractive or repulsive for symmetric and anti-symmetric configurations respectively. The entropic contribution arises from thermal fluctuations around a minimum energy configuration and is independent of the bend angle. The result thus applies to any ligand binding that has a local stiffening effect. The entropic potential is found to be always attractive in the presence of tension, meaning that the proteins bound to the DNA effectively attract each other. We would like to note that to obtain the entropy the authors use a theorem proposed by Coleman [167], however the boundary conditions required in the theorem are not fulfilled. Hence the Rudnick and Bruinsma result requires a revision. If valid, these finding have implications particularly for one-dimensional diffusion of proteins along the DNA.

Diamant and Andelman study the effect of coating DNA with small molecules such as surfactants, lipids or folded proteins [165]. Due to the relatively large length scales considered ($1 \ll l_p \ll N$, where $l_p \approx 500\,\text{Å}$ is the DNA persistence length and N the number of monomers) the cooperativity can be investigated with the use of mean field theory. The authors assume that the large number of binding molecules alter the effective

persistence length of the DNA molecule. The authors divide the problem into the two limiting cases, weak and strong coupling. In the weak coupling limit, in which the total persistence length of DNA is modified only negligibly and which is thus closer to the system studied in this thesis, the binding is found to be anti-cooperative. Significant cooperativity is observed only when the stiffness of the chain is weakened or increased by a large enough factor of about 5.

Rappaport and Rabin investigate the influence of cooperative binding of DNA-deforming proteins on the on the local spontaneous curvature of DNA [166]. This question is not considered in this thesis.

Dynamic allostery in DNA has been observed and measured directly on a small artificial ligand called Hoechst33258 cooperatively binding to a short oligomer [101, 155]. H33258 is an A/T selective, minor groove binding compound with a potential for anti-cancer applications [168]. Searle and Embrey constructed a DNA dodecamer d(CTTTTGCAAA-AG)$_2$ with two potential H33258 binding sites separated by two GC base pairs (Fig. 4.1). They found that H33258 binds to the oligomer in a highly cooperative manner, with only doubly liganded (2:1) complexes being observed in equilibrium with the free DNA [169]. The solution NMR structure of the complex shows that the two H33258 molecules bind symmetrically into the minor groove of the 5'-TTTT and 5'-AAAA sites. The NMR
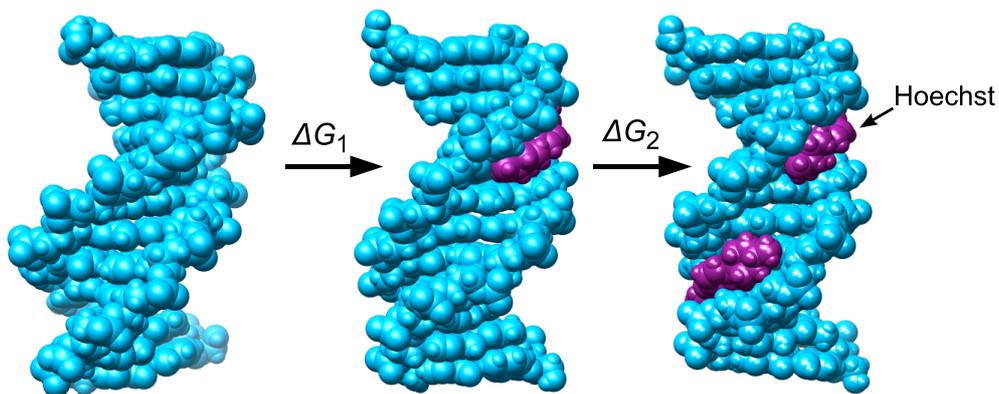
FIGURE 4.1: Binding of the drug Hoechst33258 (purple) to the DNA oligomer (blue). The free energies calculated are shown, $\Delta\Delta G = \Delta G_2 - \Delta G_1$. The figure was extracted from the simulations of Harris *et al.* [101].

structure was later refined by Gavathiotis *et al.* [155] who searched for the origins of the cooperativity. They confirmed that the two ligands are separated by the GC basepairs and are oriented so that their positively charged piperidine rings face each. The charged centers are separated by ~15 Å and the intervening water molecules are believed to screen these charges. This excluded the possibility of the signal propagation through a direct contact between the ligands. The structure otherwise gave no indications for the origins of the cooperativity, in particular the structure of the DNA in the 2:1 complex was nearly identical to the free oligomer. The minimum allosteric free energy $\Delta\Delta G$ was estimated to equal -4 kcal/mol = -6.8 k_BT.

Hamdan *et al.* [170] studied a different dodecamer d(AATTATATTAT)$_2$ and observed H33258 and another minor groove binding ligand, netropsin, to also bind with strong cooperativity. Since capillary electrophoresis, the technique used in their experiments, is not suited for structure determination and because these complexes have not been studied since, details of the interaction remain unknown. It is probable however that H33258 and structurally closely related netropsin use the same allosteric mechanism.

Harris *et al.* [101] used molecular dynamics to study the interaction of H33258 with the dodecamer of Searle and Embrey. They constructed the experimentally unobserved 1:1 complex and confirmed the hypothesis that binding of the first drug causes a negligible conformational change in the binding region of the second drug (for snapshots from their simulations see Fig. 4.1). In addition they calculated thermodynamic and kinetic parameters of the allosteric reaction. Their estimate of the allosteric free energy $\Delta\Delta G = -11.8 \pm 1.3\,k_B T$ is much larger in value than the experimental result. The entropic contribution to the allosteric free energy was obtained using the Schlitter version of quasiharmonic analysis (see section 2.2.2.1 or reference [95]), and it was found that the cooperativity is entropy driven; $T\Delta\Delta S$ was estimated at $17.5 \pm 1.2\,k_B T$ at T = 300 K. Despite their simulation being relatively short (5 ns), the thermodynamic parameters converged well.

Laughton with his coworkers repeated the simulations with the most recent force field parmBSC0 and calculated the free energy and entropy from 10 ns simulations. The results agree qualitatively with their previous calculation but are in general of larger values $\Delta\Delta G = -22\,k_B T$, $\Delta\Delta H = 5\,k_B T$ and $T\Delta\Delta S = 27\,k_B T$.

In their simulations Harris *et al.* observed an extensive change in the dynamics of the DNA; the flexibility of the DNA molecule was dramatically reduced in the occupied

system	$(TS)_{top}$ / kcal·mol^{-1}	$(TS)_{bottom}$ / kcal·mol^{-1}
free DNA	366.3	368.4
1:1 complex	341.7	363.8
2:1 complex	348.1	346.9

TABLE 4.1: Configurational entropies for bottom and top halves of the DNA oligomer extracted from simulations of Harris *et al.* [101]. DNA oligomer with one molecule of H33258 bound is denoted as 1:1 complex and with two molecules bound as 2:1. The entropy is markedly reduced in the half with Hoechst 33258 is attached (top half in the 1:1 complex).

A-tract but also noticeably reduced in the unoccupied part. This can be seen from configurational entropy for each half of the DNA molecule (see Table 4.1) and from the minor groove width measurements shown in Fig. 4.2. They concluded from the normal mode analysis that the reduction in configurational entropy is not a result of suppressing one or a few modes, but a net result of a small contribution by each mode.

Rahimian *et al.* [171] conducted melting assays and NMR experiments that confirmed a strong cooperativity of the DNA oligomer binding H33258 and another minor groove binding compound called DB183. They demonstrated that when a short loop was added at the end of the duplex to create a hairpin structure of the DNA the cooperativity was significantly reduced in both cases. This is believed to be due to reduced fluctuations of the hairpin compared to the duplex, resulting in less cooperative dynamic allostery.

These results and physical intuition suggest that "allostery without conformational change", proposed by Cooper and Dryden [18] and applied to several diverse classes of proteins [38, 41, 42, 126], plays an important role in DNA too. In the first approximation DNA behaves like an inextensible elastic rod that can twist and bend [98, 172]. The aim of this chapter is to evaluate the allosteric signal obtained from local perturbation of the twisting and bending rigidities caused by the ligand binding. The results are analysed for a general system and then parameterised for the case of Hoechst association with DNA.
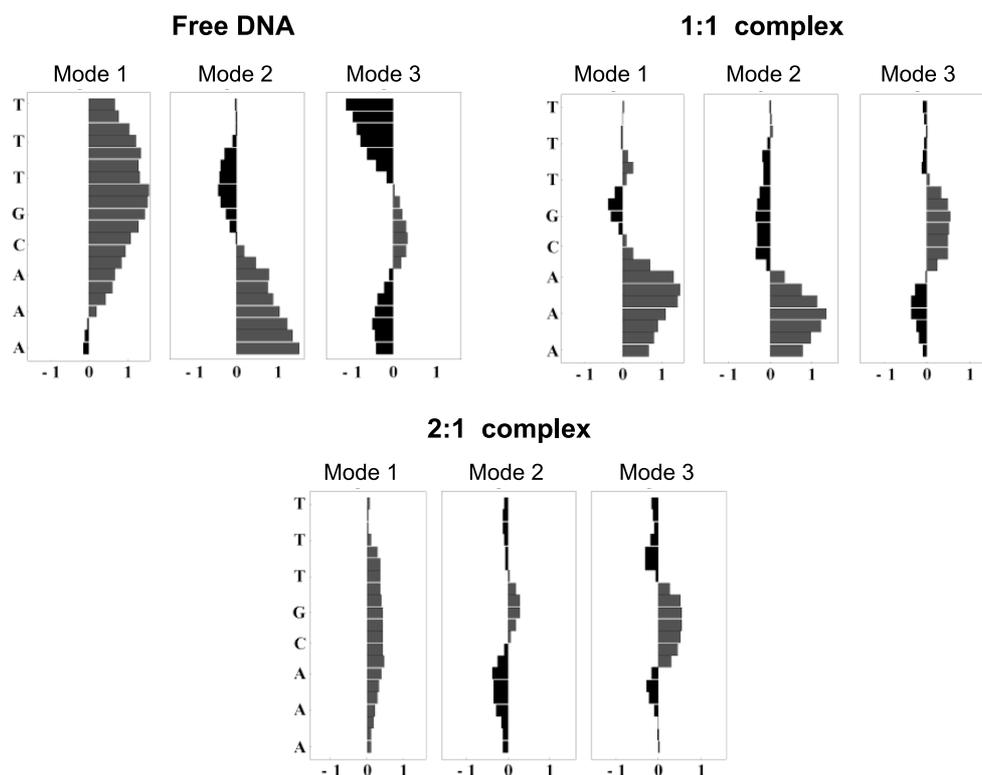
FIGURE 4.2: Minor groove width deviations (in angstroms) along the helical axis associated with first three normal modes. The deviations were obtained by subtracting the groove widths at the extremal points of each normal mode. Taken from [101].

4.2   Elastic Rod Model

4.2.1   Approximation of DNA as an Elastic Rod

DNA is assembled from two anti-parallel linear polymer chains held together by hydrogen bonds. The details of the interaction between the two chains and the secondary structure of DNA was discovered by Watson and Crick [173]. Each chain, or strand, is composed of a sequence of nucleotides formed by a sugar, a base and a phosphate (see Fig. 4.3). There are four types of bases, adenine (A), guanine (G), thymine (T) and cytosine (C). The bases are situated at the interface of the two strands, they are joined in pairs and held together by hydrogen bonds. The pairing is specific, only A/T and G/C pairs form bonds. Importantly the two polynucleotide chains are wound around a common axis to form a double helix (Fig. 4.3). The diameter of the helix of a B-DNA molecule, the most commonly observed DNA structure in nature, is approximately 20 Å, the adjacent bases
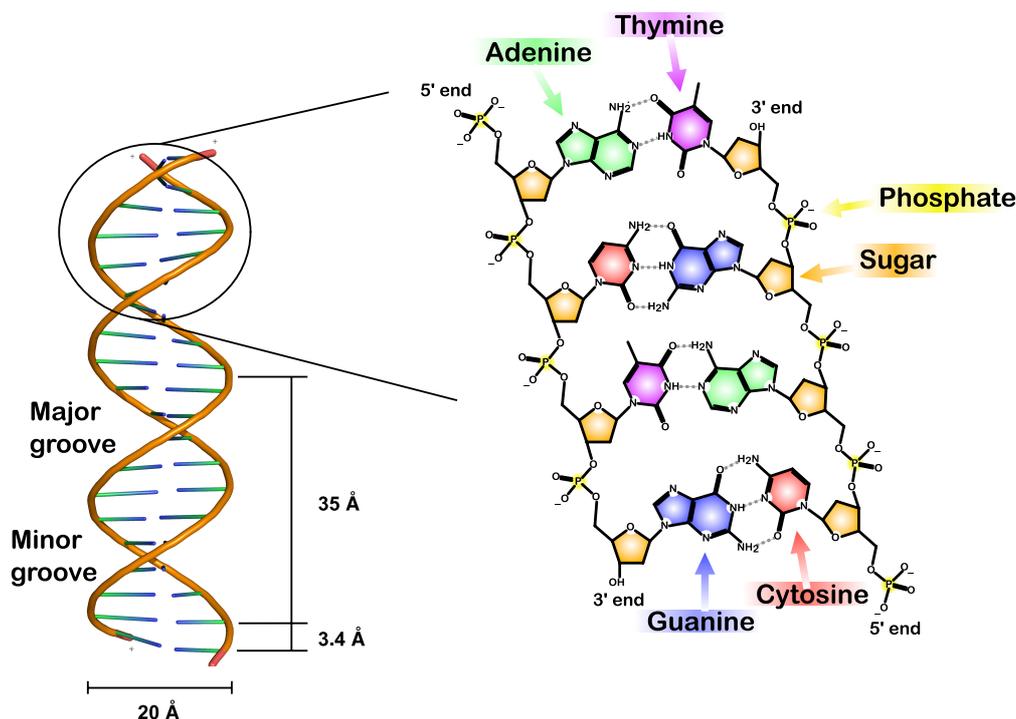
FIGURE 4.3: Left: DNA double helix shown in the cartoon representation. The backbone is shown as an orange tube. Right: Detail of the chemical structure of DNA.

are 3.4 Å apart and are rotated by ~ 35° with respect to one another [174–176]. Another particularity of the DNA structure is that the strands do not lie directly opposite each other, which results in unevenly sizes clefts at the surface of the helix. These are called the major and minor grooves and often act as binding sites for small ligands or proteins.

Although the chemical structure of DNA is indisputably important for ligand binding, we believe that also the dynamic properties of the DNA play a crucial role. In a thermal environment DNA undergoes a whole range of internal motions ranging from oscillations of individual atoms (picosecond frequencies), to processes that take seconds such as writhing or isomerisation. For a classification of the motions see Table 4.2. Allosteric effects are propagated across long distances mainly by global, low frequency motions. Capturing these global motions is possible with coarse-grained models, without necessarily the need for microscopic detail. As proposed by Barkley and Zimm [172], the appropriate description of the global motions of the DNA molecule is provided by the worm like chain (WLC) model. The validity of this description around equilibrium has been shown by several experimental [177–180], computational [176, 181, 182] and theoretical studies [172, 183–185].

	Picosecond	Nanosecond	Microsecond	Milisecond	Second
Types of motion	Atom oscillations	Oscillations of small groups of atoms; bending and twisting motions of the double helix	Bending motions winding and unwinding of the double helix; opening of base pairs	Dissociation of the double helix (unwinding)	Writhing, isomerization, division of bacteria
Amplitudes	10^{-1} Å	$5 - 7$ Å			$2 - 3\,\mu$m

TABLE 4.2: Classification of DNA motions. Adapted from [154].

Importantly, the WLC model accounts for DNA's appearance in water solution, where it looks like a rod or a random coil depending on the size of the molecule. Moreover, Barkley and Zimm demonstrated that the model explains rotational oscillatory motions of the DNA molecule observed in fluorescence depolarization experiments [172].

With the development of experimental techniques, in particular single molecule manipulation, more precise measurements of DNA mechanical properties that offer a test to the WLC model became possible. Smith *et al.* used magnetic beads in force extension experiments to show that in the low force regime DNA behaves as an entropic spring but also that the data cannot be fitted with the freely jointed chain model (FJC) [177]. FJC is the simplest model of elastic properties of a polymer that nevertheless describes well the behaviour of the majority of synthetic polymers and unfolded proteins [186]. Bustamante *et al.* demonstrated that force extension experiments can be fitted with the WLC model with excellent precision (see Fig. 4.4) [183]. The groups of Bustamante and Block later showed that even in the high force limit the DNA obeys the WLC model [178, 179]. However, the model needs to account for the extensibility of the molecule in high force regime by including a stretching term in the constitutive equation [187].

The basic force extension experiments provide a good estimate of the overall persistence length of the DNA: l_b = 53 nm [180]. The persistence length l_b will be precisely defined in Sec. 4.2.2, for a moment let us provide a qualitative explanation. It sets a measure for the bending properties of the molecule in the thermal environment; if the radius of curvature is much longer than the persistence length then the energy required to bend the rod is negligible compared to the thermal energy $k_B T$. Alternatively, directional correlations along the polymer are significant for segments separated by less than a persistence length, but negligible beyond it.
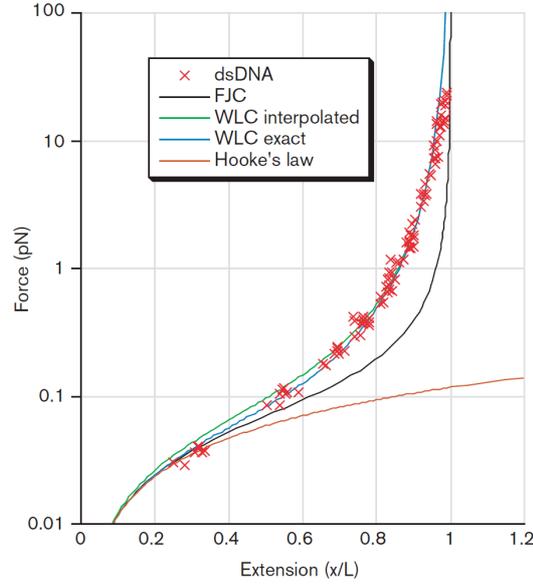
FIGURE 4.4: The force extension data for λ phage DNA of Smith *et al.* [177] fitted to the FJC and WLC model solved numerically (WLC exact) and to an interpolation expression (WLC interpolated). The WLC provides an excellent fit for low and intermediate forces. For lowest forces DNA behaves like a Hookian spring, however the Hooke's law as well as the FJC model fail to fit the data for intermediate forces. Adapted from [180].

Twisting rigidities were first obtained with good precision from experiments of Strick *et al.*, who devised an improvement of the extension experiments to allow for the torsional constraints to be applied. They confirmed the estimates of the elasticity theory predicting the twist persistence length to be $\approx 75\,\mathrm{nm}$ [188]. The twist persistence length is defined analogously to the bend persistence length.

The force extension experiments provide arguably the most accurate test to the WLC model. Other experimental techniques used for the investigation of mechanical properties of DNA include cyclisation experiments [82, 189], electron microscopy [190] and electron paramagnetic resonance [191]. Furthermore, computational methods offer a complementary tool to test the WLC model. Molecular dynamics and Monte Carlo simulations have been employed on many occasions to test or refine the experimental findings [176, 181, 192–195].

In particular, computational methods are suitable for studying the sequence dependence of the rigidities. It is well known that the sequence in one region affects physical properties of the surrounding regions [20, 196]. However, experiments generally measure the

elastic properties of DNA on large molecules where the local variations cancel. Computer simulations are on the other hand limited in size, and therefore more suited for studying short sequences in much more detail. The most accepted results on the sequence dependent DNA flexibility have been obtained from molecular dynamics simulations, e.g. [176, 197]. The sequence dependency of the parameters has also been obtained from analysing x-ray and NMR structures of all available high resolution DNA-protein complexes [176, 198]. More details on the methods used to evaluate the sequence dependent rigidities will be given in Sec. 4.2.3.4.

In the following sections we evaluate the dynamical allostery in an isotropic elastic rod like molecule. This is the right starting point since DNA can be modeled as an isotropic elastic rod on the medium-large scale, but we can already anticipate that inhomogeneity may emerge when shorter segments are studied in more detail. In Sec. 4.2.3.4 we extract the elastic constants from the simulations of Harris *et al.* and show that despite the constants slightly varying along the oligomer the effect of the ligand binding is much larger. Thus inhomogeneity is neglected throughout all our calculations.

4.2.2   Methods

In the wormlike chain model the molecule is treated as a thin, continuously flexible elastic rod of length L. The rod is immersed in viscous fluid at thermal equilibrium. This system is well studied in physics and the results for bending, twisting and stretching motions are known [199]. In the first approximation the stretching can be neglected due to its comparably high modulus. The elastic energy of the rod is then composed of two parts, the bending and the twisting energy $E = E_b + E_t$. The individual contributions have been derived from continuum mechanics in Appendix B and read

$$E_b = \frac{1}{2} \int_0^L A(s) \left(\frac{\partial^2 \mathbf{r}}{\partial s^2} \right)^2 \mathrm{d}s$$

$$E_t = \frac{1}{2} \int_0^L C(s) \left(\frac{\partial \theta}{\partial s} - \Omega_0 \right)^2 \mathrm{d}s, \tag{4.1}$$

where $\mathbf{r}(s)$ is the position vector, $\theta(s)$ is the twist angle, A and C are the bending and twisting rigidity respectively that are in principle functions of s, the arc length. A and C are related to the bend and twist persistence lengths as $l_b = A/k_B T$ and $l_t = C/k_B T$. We
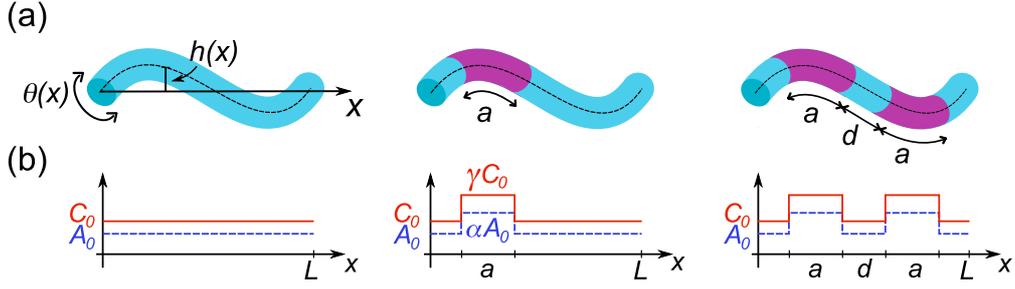
FIGURE 4.5: (a) Schematic representation of the elastic deformation of a thin elastic rod. The parameter $h(x)$ is the deflection of the axis of the rod from its relaxed straight state, $\theta(x)$ is the twist angle at the position x. The location of ligand binding is indicated by purple stripes. The geometry of the system is described by L, the length of the rod, a the length and d the separation of the ligands. (b) The influence of the binding on the bending and twisting rigidity. The bending rigidity A_0 is modified by a factor α and the twisting rigidity C_0 by factor γ in the location of binding.

are interested in free energy differences. The logarithmic dependence of the free energy on the elasticity matrix (see Eqs. (2.21) and (2.23)) converts the differences of matrix determinants into their ratios and the absolute values of the parameters A and C cancel. We define the coordinate system so that the x-axis is parallel to the axis of the straight undeformed rod. The position vector can then be written as $\mathbf{r}(s) = [x, h(x)]$, where $h(x)$ is the deviation of the rod from the relaxed, straight state. A schematic representation of the parameters is shown in Fig. 4.5. For small deviations from the relaxed state the bending energy can be written as

$$E_b = \frac{1}{2} \int_0^L A(x) \left(\frac{\partial^2 h}{\partial x^2} \right)^2 \mathrm{d}x. \tag{4.2}$$

The energy of the torsional deformation can be written similarly in terms of x. Because the DNA molecule is twisted in a relaxed state, the relaxed twist density Ω_0 is included in the energy. The accepted value of 10.5 base pairs per helical turn [176, 186] results in the value of

$$\Omega_0 = \frac{2\pi}{10.5\,\mathrm{bp}} \frac{1\,\mathrm{bp}}{0.34\,\mathrm{nm}} \sim 1.8\,\mathrm{nm}^{-1} \tag{4.3}$$

We assume that Ω_0 is not modified upon the ligand binding and consequently does not contribute to the allosteric free energy. The constant term can therefore be left out from Eq. (4.1) in the rest of the chapter.

In order to evaluate the contribution of the fluctuations to the free energy we need to

find the eigenfunctions of the Hamiltonians Eq. (4.1) and (4.2). This requires converting the energy equations into a general form

$$E = \int_0^L f \mathcal{L} f \, \mathrm{d}x, \tag{4.4}$$

where $\mathcal{L}$ is a differential operator and f is either h or θ for bending and twisting respectively. By solving the eigenvalue problem for the operator $\mathcal{L}$ we obtain the eigenfunctions of the respective Hamiltonians. Appendix C shows the derivation of the eigenfunctions for A = const. and C = const. The deviation of the rod and the twist angle can be written as a linear combination of these normal modes

$$
\begin{aligned}
h(x) &= \sum_{n=0}^{\infty} a_n h_n(x) \\
\theta(x) &= \sum_{n=0}^{\infty} b_n \theta_n(x),
\end{aligned}
\tag{4.5}
$$

where a_n and b_n are the amplitudes of the n-th mode. The eigenmodes for the free end boundary conditions (Eqs. (C.4) and (C.14)) read

$$
\begin{aligned}
h_n(x) &= g \left[(\cos(\sigma_n L) - \cosh(\sigma_n L))(\cos(\sigma_n x) + \cosh(\sigma_n x)) \right. \\
&\left. + (\sin(\sigma_n L) + \sinh(\sigma_n L))(\sin(\sigma_n x) + \sinh(\sigma_n x)) \right],
\end{aligned}
\tag{4.6}
$$

$$\theta_n(x) = \sqrt{\frac{2}{L}} \cos\left(\frac{n\pi x}{L}\right), \quad n = 1, 2, \dots \tag{4.7}$$

σ_n is the n-th root of the equation

$$\cosh(\sigma L)\cos(\sigma L) = 1. \tag{4.8}$$

This equation can only be solved numerically and the first few solutions are

$$\sigma_n L \doteq 4.73, 7.853, 10.996, \dots \tag{4.9}$$

The solutions can be approximated by [199]

$$\sigma_n L \doteq \frac{(2n+1)\pi}{2}. \tag{4.10}$$

for higher n. This approximation is very good yielding the first few values of $\sigma_n L$ = $4.71, 7.854$ and 10.996. The two sequences converge with increasing n.

The functions h_n become very hard to manipulate with increasing n, because of the exponential divergence of the hyperbolic functions. We approximate h_n for large n so that

$$h_n(x) \doteq \sqrt{\frac{\sigma_n L}{\sigma_n L - 1 - \cos^2(\sigma_n L)}}(\cos(\sigma_n x) - \sin(\sigma_n x)), \quad n \gg 1, \qquad (4.11)$$

The detailed derivation is presented in Appendix C. The comparison of the original and approximated expressions for first few n are shown in Fig. 4.6.

Finally, in order to dispose of the length dependence we introduce a new rescaled variable

$$\hat{x} = \frac{x}{L}. \qquad (4.12)$$
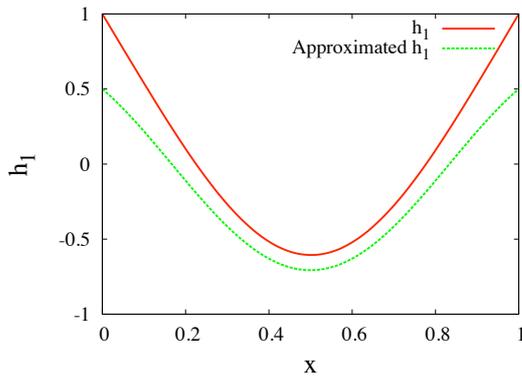
(a) $n = 1$ (b) $n = 2$
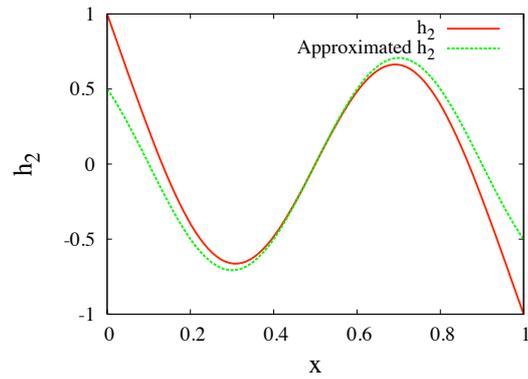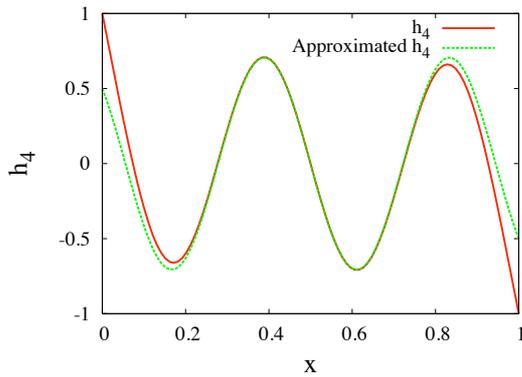
(c) $n = 4$ (d) $n = 10$
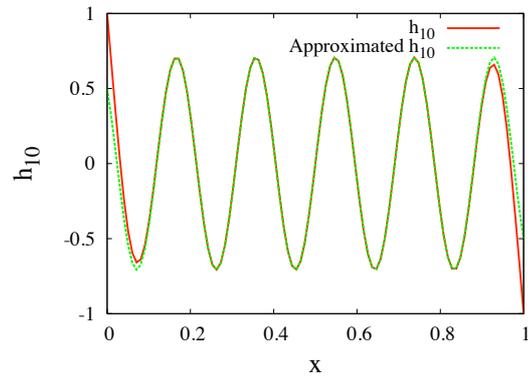
FIGURE 4.6: Comparison of exact and approximated normal bending modes for $n = 1, 2, 4$ and 10.

All subsequent calculations are presented in terms of $\hat{x}$. We substitute $\hat{x}$ into the expressions for energy (4.1) to obtain

$$E_b = \frac{1}{2} \int_0^1 \tilde{A}(\hat{x}) \left(\frac{\partial^2 h(\hat{x})}{\partial \hat{x}^2} \right)^2 \mathrm{d}\hat{x} \tag{4.13}$$

$$E_t = \frac{1}{2} \int_0^1 \tilde{C}(\hat{x}) \left(\frac{\partial \theta(\hat{x})}{\partial \hat{x}} \right)^2 \mathrm{d}\hat{x}, \tag{4.14}$$

where $\tilde{A} = A/L^3$ and $\tilde{C} = C/L$. The substitution has to be also carried out in the expressions for normal modes Eq. (4.6), (4.7) and (4.11).

Any molecule binding to the elastic rod is expected to locally modify the flexibility of the rod. Local stiffening and softening of DNA as a result of binding proteins has been observed in several studies [200–202]. Over the next sections we calculate propagation of the signal activated by the rigidity modification along the elastic rod. We aim to evaluate to what degree the change in flexibility at one location can facilitate or hinder the binding of another molecule to a different segment of the rod. In order to reduce the number of free parameters in our model we assume that the second molecule is identical or has the identical effect locally on the flexibility of the elastic rod as the first molecule. Different local effects can be introduced into the model if needed. We model the local change of flexibility as a local change of the rigidity constants. Two non-dimensional parameters are introduced to describe the effect of the binding; in the location of the ligand binding the bending rigidity A changes by a factor α and the twisting rigidity C by a factor γ.

The geometry of the system is shown in Fig. 4.5. The two ligands bind a distance d apart and each of them occupies the length a of the unscaled rod. The change of the bending and twisting rigidity evoked by the interaction is illustrated in the same figure.

The decomposition of the rod deviation h and the twist angle θ into the normal modes leads to a convenient evaluation of the energy and subsequently of the partition function. For example, inserting the decomposition of the twist angle Eq. (4.5) into the energy term written in the form of (4.4) for the free, homogeneous rod yields

$$E_t = \int_0^1 \sum_{n,m=1}^{\infty} b_n b_m \theta_n \mathcal{L} \theta_m \mathrm{d}x = \int_0^1 \sum_{n=1}^{\infty} b_n^2 \lambda_n \theta_n^2, \tag{4.15}$$

where λ_m is the eigenvalue associated with the $m-$th eigenfunction. The eigenfunctions are orthogonal which reduces the summation to one index. The energy can equivalently be written in a matrix form

$$E_t = \frac{1}{2}\mathbf{b}^{\mathrm{T}} M_t \mathbf{b}, \tag{4.16}$$

where M_t is a diagonal matrix of the eigenvalues λ_n. The bending energy is expressed in the same way with either approximated or numerically calculated eigenvalues on the diagonal. The differential operator can be in principle found for each binding state and the corresponding differential equations solved to yield piecewise defined eigenfunctions. However, we use an alternative method.

The functions h_n and θ_n derived for the special case of homogeneous rod provide a basis of the real vector space. We can therefore calculate the energy matrices for the singly and doubly liganded geometries by inserting the decompositions (4.5) into the modified integrals.

Thus for example the energy of twisting of a singly liganded rod (see Fig. 4.5 for the geometries)

$$E_t = \frac{1}{2}\mathbf{b}^{\mathrm{T}} M_{t,1:1} \mathbf{b}. \tag{4.17}$$

and the matrix is defined by

$$[M_{t,1:1}]_{i,j} = \left(\int_0^{l_1} + \gamma \int_{l_1}^{l_2} + \int_{l_2}^{1} \right) \frac{C_0}{2} \left(\frac{\partial \theta_i(\hat{x})}{\partial \hat{x}} \right) \left(\frac{\partial \theta_j(\hat{x})}{\partial \hat{x}} \right) \mathrm{d}\hat{x}, \tag{4.18}$$

where the distances $l_1 = (L - d - 2a)/(2L)$ and $l_2 = (L - d)/(2L)$ are found from Fig. 4.5. Similarly the energy of the 2:1 complex consists of five integrals. Similar results are obtained for different binding states (geometries) and for bending.

The partition function of a particular binding state is calculated as

$$Z = \iint_{\mathbb{R}^{2N}} \mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b}\,\exp\left(-\frac{\mathbf{a}^{\mathrm{T}} M_b \mathbf{a} + \mathbf{b}^{\mathrm{T}} M_t \mathbf{b}}{2k_B T} \right). \tag{4.19}$$

This is a multidimensional Gaussian integral that can be evaluated to give

$$Z = \left(\frac{(2\pi k_B T)^n}{|M_b|} \right)^{\frac{1}{2}} \left(\frac{(2\pi k_B T)^n}{|M_t|} \right)^{\frac{1}{2}}, \tag{4.20}$$

where $|M| = \det M$. The free energy is then

$$G = -k_B T \ln Z = -\frac{1}{2} k_B T \ln \left(\frac{(2\pi k_B T)^n}{|M_b|} \right) - \frac{1}{2} k_B T \ln \left(\frac{(2\pi k_B T)^n}{|M_t|} \right) \qquad (4.21)$$

We are interested in the allosteric free energy which is the difference in the free energy change between the two binding steps $\Delta\Delta G = \Delta G_2 - \Delta G_1 = (G_{2:1} - G_{1:1}) - (G_{1:1} - G_{free})$. Inserting Eq. (4.21) into this expression and rearranging gives

$$\Delta\Delta G = \frac{1}{2} k_B T \left(\ln \frac{|M_{b,free}||M_{b,2:1}|}{|M_{b,1:1}|^2} + \ln \frac{|M_{t,free}||M_{t,2:1}|}{|M_{t,1:1}|^2} \right) \qquad (4.22)$$

Thus the total allosteric free energy is a sum of individual contributions from twisting and bending.

This method bypasses the generally difficult evaluation of the true eigenmodes of the DNA in the bound, inhomogeneous state. On the other hand, it introduces a complication to the determinant evaluation in the biologically relevant case of finite number of modes.

Physically, only a finite number of modes can be excited in a DNA, or any other rod-like molecule, due to the finite number of atoms. Mathematically, this means substituting the determinant of the infinite matrix in Eq. (4.21) by a product of the first n eigenvalues. But the true eigenvalues were found only for the homogeneous rod. The set of the corresponding eigenfunctions was then used to express the Hamiltonians of the singly and doubly liganded rod. The determinant of the matrix in this basis, truncated at the n-th basis function, is then not equivalent to the product of the first n true eigenvalues of the respective Hamiltonian.

Fortunately, a majority of eigenvalues of this truncated matrix is nearly identical to the true eigenvalues. Moreover, the "correct" eigenvalues are the lowest eigenvalues and the erroneous eigenvalues can be easily identified in the eigenvalue plot (Fig. 4.7). In the allosteric free energy calculations the erroneous eigenvalues are disregarded and the determinants substituted by the product of the first n correct eigenvalues. In Sec. 4.2.3 we evaluate the true eigenmodes for a special choice of parameters and show that the resulting free energies are identical with those obtained with the method introduced above.
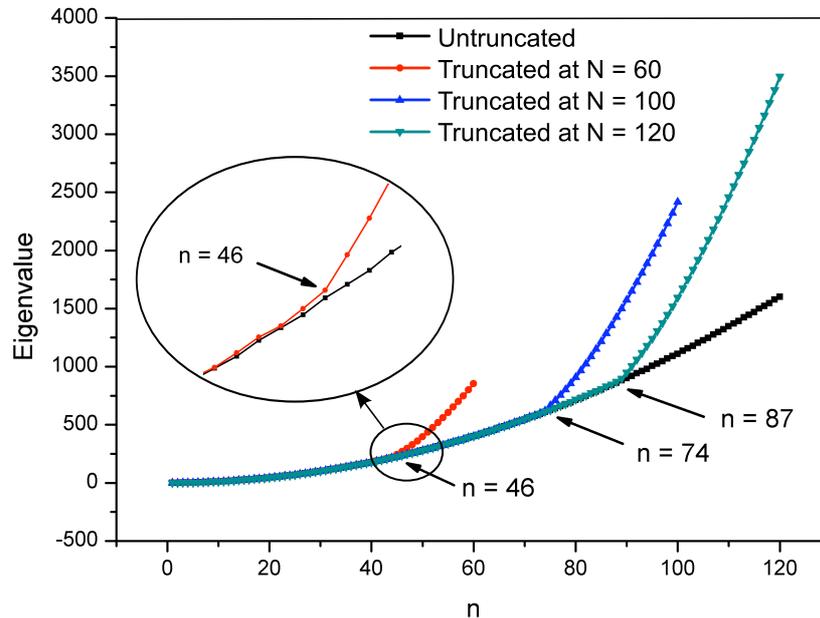
FIGURE 4.7: Eigenvalues of the untruncated matrix (black) and the matrix truncated at several points (N = 60, 100 and 120). The truncated matrix eigenvalues agree well with the explicit calculations until a "kink" whose location occurs at ~ 2/3N. n is the mode number, $n = 1, \ldots, N$.

4.2.3   Results

Let us first evaluate the theoretical contribution to the allosteric free energy resulting from the perturbation of the twisting modes. The calculation is carried out for a general geometry and values of parameters using the method described above. The allosteric free energy is found always to inhibit the second binding irrespective of the geometry or parameters. In order to understand this result we evaluate the normal modes for a special choice of parameters and find that the negative cooperativity arises from increased amplitude of fluctuation in the unliganded part of the elastic rod. In the second part of this section we focus on the bending motions and find a result analogous to that of the twisting. Thorough analysis with respect to the free parameters is presented.

4.2.3.1 Twisting Deformations

The energy of the free elastic rod is

$$E_{t,free} = \int\limits_0^1 \frac{C_0}{2} \left(\frac{\partial \theta(\hat{x})}{\partial \hat{x}} \right)^2 \mathrm{d}\hat{x}, \tag{4.23}$$

where C_0 is a constant. The energy matrix M_t (Eq. (4.16)) is diagonal with the eigenvalues

$$\lambda_n = \frac{C_0}{2} \pi^2 n^2. \tag{4.24}$$

When the first and second ligand bind the rigidity $\tilde{C}(\hat{x})$ becomes a rectangular function as shown in Fig. 4.5. The energy matrix of the singly bound rod is obtained from Eq. (4.18). Similarly the energy of the 2:1 complex consists of 5 integrals.

As explained in Sec. 4.2.2 only a finite number of modes can be excited in a DNA, or any other rod-like molecule, due to the finite number of atoms. We introduce a cutoff by relating the wavelength of the shortest mode to a physically relevant distance on the DNA molecule. Two characteristic distances are used: 1) the distance between neighboring base pairs and 2) the distance between the atoms on the backbone of the DNA. In a DNA molecule there are 6 atoms per base pair (Fig. 4.3) and thus for an oligomer with i base pairs we evaluate the free energy for i and $6i$ normal modes to set a range for the expected contribution of thermal fluctuations to the allosteric effect.

The determinants of the energy matrices (Eq. 4.22) are substituted by a product of the desired number of eigenvalues (e.g. i or $6i$ in the case of DNA oligomer with i base pairs). Let us compare the eigenvalues of the truncated basis set matrix to the true eigenvalues. We find a special set of parameters ($a = 1/2$, $d = 0$ and $\gamma = 4$) for which the eigenvalues can be obtained analytically. Justification of the parameter choice and details of the calculation will be given in the next section. The plot of the exact eigenvalues overlaid with the eigenvalues obtained for the same parameters from the truncated matrix (Fig. 4.7) reveals a sharp "kink" in the approximate eigenvalue sequence. The approximate eigenvalues with a mode number smaller than the kink are nearly identical to the exact ones, but for mode numbers larger they differ substantially.

We assume that the eigenvalues below the kink are always nearly exact and that we can calculate the determinant of the truncated matrix as a product of the eigenvalues

lower than the kink location. This method could be potentially used to extrapolate the determinant of the infinite matrix, but because the cutoff has a physical meaning we only use the values for the truncated matrix. A more detailed analysis of the eigenvalue plots is presented in Appendix D.

$\Delta\Delta G$ has been evaluated for all combinations of parameters and was found to be always positive, corresponding to negative cooperativity. In other words binding of a molecule to the elastic rod and thus modifying its twisting rigidity results in reduced affinity for the binding of the second ligand to a different segment of the rod.

As shown in Fig. 4.8 the allosteric free energy grows linearly with the number of modes N included in the system. The oscillations at low N arise from the truncation and become smoothened out as the size of the truncated matrix increases. The allosteric free energy becomes larger, or the system more anti-cooperative, as the enhancement factor γ increases. The dependency on γ is nearly logarithmic for $\gamma > 1$. The logarithmic function is expected from an allosteric system undergoing subsequent tightening as has been shown by Hawkins and McLeish [38] and explained in Sec. 2.5.

The allosteric free energy as a function of the geometry of the system is shown in Fig. 4.8 c,d. As expected $\Delta\Delta G$ increases with the length of the ligand a and displays a power law dependence with the exponent becoming larger with the value of γ. The allosteric free energy in our model is independent of the distance between the ligands (the slight fluctuations observed in Fig. 4.8 are a consequence of the numerical approximations). This result is consistent with earlier calculations of Rudnick and Bruinsma who found that there is no effective interaction between the two ligands in the absence of tension, i.e. the allosteric free energy is independent of the distance between the ligands [164]. Rudnick and Bruinsma, however, did not evaluate the allosteric free energy in the sense considered here and only focused on the effective interaction that arises in the presence of tension. Our calculations address a slightly different phenomenon but are also more detailed and thorough than their work.

One can argue that the independence of $\Delta\Delta G$ on the distance between ligands is unrealistic. The anti-cooperative effect reduces to zero in the large ligand spacing limit in the real DNA. The unvanishing effect observed in our approximation is a consequence of the original assumption of elastic homogeneity of the rod and the consequent spatial delocalisation of the normal modes.

(a)

(b)

(c)

(d)
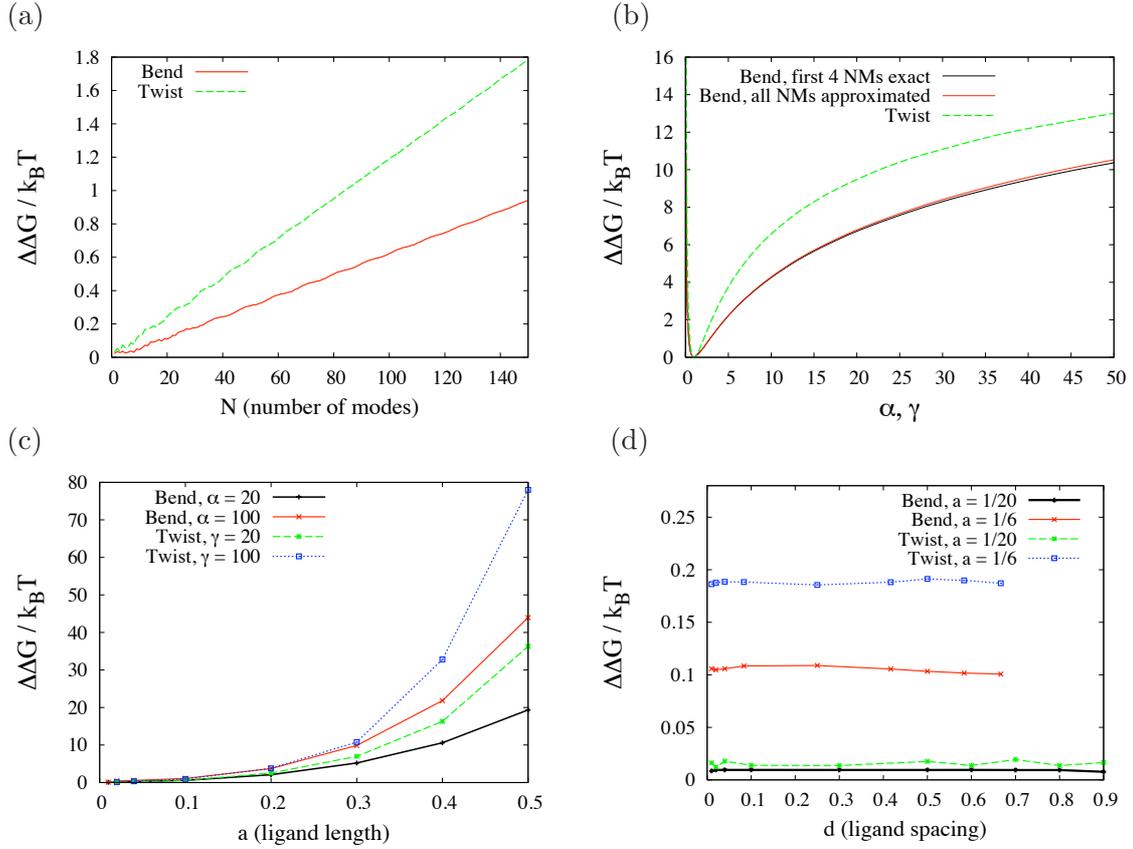
FIGURE 4.8: The allosteric free energy associated with bending and twisting as a function of (a) number of modes, (b) the rigidity enhancement factors α and γ, (c) the ligand length and (d) the distance between the ligands. For the graph (a) the matrix was truncated to dimension 250×250 and the first N eigenvalues were used to calculate $\Delta\Delta G$. (b) The dependence on the enhancement factors α and γ is approximately logarithmic for α, $\gamma > 1$ with the coefficient dependent on the type of motion and geometry of the system. Here the geometry is defined by $a = 1/3$, $d = 1/6$, $\alpha = 2$ $\gamma = 2$ and $N = 70$. (c) A polynomial increase of the allosteric free energy with the ligand length depends on the enhancement factors. The exponent varies between 2.1 for $\alpha = 2$ and 3.0 for $\alpha = 100$ and 2.3 for $\gamma = 2$ and 3.8 for $\gamma = 100$. The values of the allosteric free energy are independent of the spacing between the ligands as is shown in part (d).

We have examined also the case where $\gamma < 1$, i.e. where the twisting rigidity becomes softer as the ligand binds. We find that $\Delta\Delta G$ is still larger than 0. The dependency of the allosteric free energy on small γ can be seen in Fig. 4.8 b. The dependencies on a and d are very similar to the case of $\gamma > 1$ (data not shown).

In conclusion, the allosteric free energy associated with twisting is always positive. This is a little surprising; one might expect to find a regime where the system favors the second ligand binding. In order to understand what is happening on the level of normal modes

we visualize the modes and observe how they get modified upon the ligand binding. Since this is in general not simple we choose a special set of parameters for which the normal modes can be calculated analytically.

4.2.3.2 Visualisation of the Twisting Eigenmodes

We assume that the rod completely stiffens in each of its two half-lengths sequentially, or equivalently $a = 1/2$ and $d = 0$. The enhancement factor γ is chosen to equal 4. The normal modes of the homogeneous rod are simple cosines (Eq. (4.7)) and the energy matrix is evaluated as explained previously, resulting in eigenvalues proportional to n^2, where n is the mode number.

The normal modes of the half stiffened rod are defined piecewise to fulfill automatically the free ends condition

$$\left.\frac{\mathrm{d}\theta}{\mathrm{d}x}\right|_{x=0,L} = 0. \tag{4.25}$$

This gives

$$\theta_k = \begin{cases} \theta_{k_1} = \cos(k_1\hat{x}), & \hat{x} \in \left\langle 0, \frac{1}{2}\right\rangle \\ \theta_{k_2} = B\cos(k_2(1-\hat{x})), & \hat{x} \in \left\langle \frac{1}{2}, 1\right\rangle, \end{cases} \tag{4.26}$$

where θ_{k_1} and θ_{k_2} are the angle functions in the left and right hand side of the rod respectively. The parameters B, k_1 and k_2 can be found from the boundary conditions at the interface between the soft and hard part of the rod. Firstly, the dispersion relation [199] defines a condition for k_2

$$k_2 = \frac{k_1}{\sqrt{\gamma}} = \frac{k_1}{2}. \tag{4.27}$$

Two additional requirements are placed on the eigenmode on the interface between the bound and free part: 1) the continuity of the wave and 2) the equality of the torque on both sides.

$$1)\quad \theta_{k_1}\left(\frac{1}{2}\right) = \theta_{k_2}\left(\frac{1}{2}\right) \qquad\qquad 2)\quad \left.C_0\frac{\partial\theta_{k_1}}{\partial\hat{x}}\right|_{\hat{x}=\frac{1}{2}} = \left.\gamma C_0\frac{\partial\theta_{k_2}}{\partial\hat{x}}\right|_{\hat{x}=\frac{1}{2}} \tag{4.28}$$

These conditions can be solved analytically to yield three independent series of periodic solutions:

$$k_1 = 4\pi m \,, \quad m \in \mathbb{N} \tag{4.29}$$

$$k_1 = 4\pi m \pm 4\arctan(\sqrt{2})\,, \quad m \in \mathbb{N} \tag{4.30}$$

The amplitude B is determined for each case from condition 1) to give

$$B = (-1)^m \tag{4.31}$$

$$B = (-1)^{m+1} \frac{\sqrt{3}}{3}, \tag{4.32}$$

for the functions (4.29) and (4.30) respectively.

After normalization we obtain three sets of orthonormal, piecewise-defined eigenfunctions:

$$\theta_{1,n} = \begin{cases} \sqrt{2}\cos(4\pi n \hat{x}), & \hat{x} \in \left\langle 0, \frac{1}{2}\right\rangle \\ (-1)^n \sqrt{2}\cos\left(2\pi n(1-\hat{x})\right), & \hat{x} \in \left\langle \frac{1}{2}, 1\right\rangle, \end{cases} \tag{4.33}$$

$$\theta_{2,n} = \begin{cases} \sqrt{3}\cos((-4\arctan\sqrt{2} + 4\pi n)\hat{x}), & \hat{x} \in \left\langle 0, \frac{1}{2}\right\rangle \\ (-1)^{n+1} \sqrt{2}\cos\left((-2\arctan\sqrt{2} + 2\pi n)(1-\hat{x})\right), & \hat{x} \in \left\langle \frac{1}{2}, 1\right\rangle \end{cases} \tag{4.34}$$

and

$$\theta_{3,n} = \begin{cases} \sqrt{3}\cos((4\arctan\sqrt{2} + 4\pi n)\hat{x}), & \hat{x} \in \left\langle 0, \frac{1}{2}\right\rangle \\ (-1)^{n+1} \sqrt{2}\cos\left((2\arctan\sqrt{2} + 2\pi n)(1-\hat{x})\right), & \hat{x} \in \left\langle \frac{1}{2}, 1\right\rangle. \end{cases} \tag{4.35}$$

These functions diagonalise the energy matrix in the half-stiffened case. The error arising from the truncation of the matrix in the sequence of non-diagonalised basis functions is thus avoided and the exact free energy can be found. The energy matrix has a diagonal block form, each block corresponding to one solution for k_1 (Eq. (4.29), (4.30)).

$$M_{t,1:1} = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix}$$

and the eigenvalues of the individual blocks are

$$
\begin{aligned}
M_1 &\longrightarrow 8\pi^2 n^2, \qquad 1 \le n \le N \\
M_2 &\longrightarrow 8\left(\pi(n-1) - \frac{1}{\pi}\arctan(\sqrt{2})\right)^2, \qquad 1 \le n \le N \\
M_3 &\longrightarrow 8\left(\pi(n-1) + \frac{1}{\pi}\arctan(\sqrt{2})\right)^2, \qquad 1 \le n \le N
\end{aligned}
\tag{4.36}
$$

The free energy of the half stiffened rod is

$$
G_{1:1} \propto -\frac{1}{2}k_B T \ln|M_{t,1:1}|
\tag{4.37}
$$

and the allosteric free energy is calculated as a difference $\Delta\Delta G = (G_{2:1} - G_{1:1}) - (G_{1:1} - G_{free})$. From a comparison to the previously presented approach we find that the difference in the value of the allosteric free energy is negligible ($< 10^{-2}\, k_B T$). But unlike in the previous approach, the eigenmodes of the half bound case can be calculated and plotted together with the free rod eigenfunctions and the effect of ligand binding observed.

We find that the binding suppresses the amplitude of the mode in the stiffer part. The amplitude in the softer part becomes larger than that of the free rod and therefore the second binding is entropically more costly. This is demonstrated in Fig. 4.9, where the corresponding modes of the two binding states have been normalised to result in identical total elastic energy, which will be the case at equilibrium. In this ensemble, reduction of mode amplitude locally corresponds to a reduction in exploration of phase space that in turn generates the entropy change on ligand binding. Since the initial amplitude of fluctuations is now higher in the second (left in Fig. 4.9) part of the rod we have to pay a higher free energy penalty to stiffen it than the first (right) half. This is the case for all modes and consequently the resulting allosteric effect is anti-cooperative and the allosteric free energy grows with the number of modes included in the calculation as demonstrated in Fig. 4.8 a.
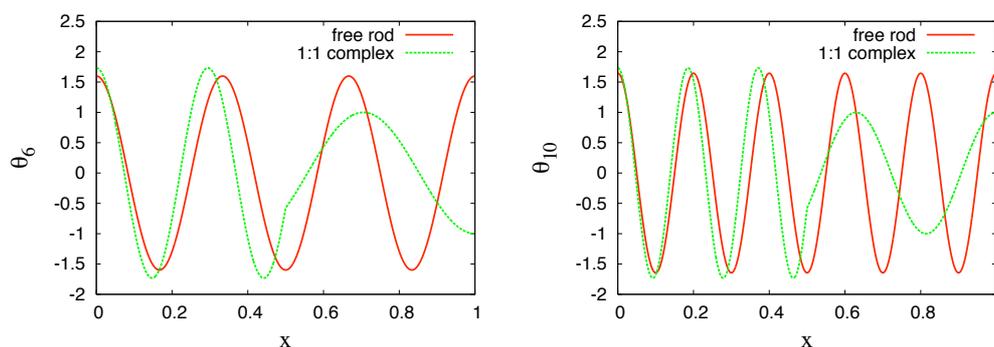
FIGURE 4.9: The 6th (left) and the 10th (right) twisting eigenmodes plotted along the rod's helical axis x for the homogeneous and half stiffened rod. The first binding occurs in the right half of the rod, the amplitude in this region is reduced but the amplitude of the left part increases. This results in entropically more costly second binding. The modes in the two binding states have been normalised to have equal elastic energy.

4.2.3.3 Bending Deformations

The elastic energy matrix of the bending deformation of the free elastic rod is diagonal with elements

$$[M_{b,free}]_{i,i} = \int\limits_0^1 \frac{A_0}{2}\left(\frac{\partial^2 h_i}{\partial \hat{x}^2}\right)^2 \mathrm{d}\hat{x}, \qquad (4.38)$$

where A_0 is a constant and $h = \sum_n a_n h_n$ is the deflection of the bent rod from its straight normal state. As for twisting, the energy of the singly and doubly liganded rod is found by integrating over the singly and doubly rectangular function $\tilde{A}$, presented in Fig. 4.5. The first four modes are described by the accurate functions Eq. (4.6) with numerically calculated values of $\sigma_n L$, for $n > 4$ the approximated functions (Eq. (4.11)) and extrapolated values of $\sigma_n L$ (Eq. (4.10)) are used.

In order to obtain physically meaningful free energies we need to introduce a cutoff to the number of modes and consequently the rank of the energy matrices. A kink occurs in the series of the eigenvalues as before (see Fig. 4.7). The determinants are accordingly calculated from the lower eigenvalues analogously to the twisting determinants.

To get an idea of the error introduced by using the approximated functions, $\Delta\Delta G$ obtained by using the precise first four normal modes and using the approximated functions only is plotted in Fig. 4.8 b. The difference becomes evident only for large α. The approximated functions converge to the accurate for higher n (fig. 4.6).

Similarly to the twisting, we analyzed the bending allosteric free energy for all possible combinations of parameters to find that $\Delta\Delta G$ is always positive. The allosteric free energy as a function of α is shown and compared to the allosteric free energy arising from twisting in Fig. 4.8 b. The character of the dependence of allosteric free energy on the other parameters (Fig. 4.8) is very similar to the twisting modes and thus the same conclusions apply.

It has to be noted that the rod can bend in two independent, perpendicular directions. Therefore every bending mode has to be counted twice, one for each polarisation. The free energies in Fig. 4.8 are presented only for one set of modes and would need to be multiplied by two for an application to a real, symmetric molecule. So assuming identical geometry and enhancement factors, i.e. $\alpha = \gamma > 1$ bending contributes more to the total allosteric free energy and the difference increases with increasing enhancement factors. For $\alpha = \gamma < 2$ however twisting becomes the slightly larger contributor.

4.2.3.4 Parameterisation of the Model for H33258 Binding DNA

We have demonstrated that modifications of bending and twisting rigidity by binding can only result in negative cooperativity. A thorough search of the literature yields however only examples of positively cooperative systems. This does not necessarily imply the absence of the negative cooperativity in nature, and we speculate on the possible reasons for the lack of experimental evidence for negative cooperativity in the conclusions section.

Here we focus on one such positively cooperative system, the artificial ligand Hoechst33258 binding to a short DNA oligomer, and assess the extent of the negative cooperativity. We show that the resulting value of $\Delta\Delta G$ is small enough to be overridden by other mechanisms associated with more subtle, DNA specific structural dynamics.

We parameterise our model from the MD simulations of the Hoechst33258/DNA system performed by Harris *et al.* [101]. We use their recent repeat of the simulations with the most recent force field parmBSC0 [203], built into the molecular dynamics software package Amber [85]. A 10 ns simulation was run for each of the three binding states. We then employ the so called base pair step model to extract twist and bend rigidities as a function of the oligomer's binding state.
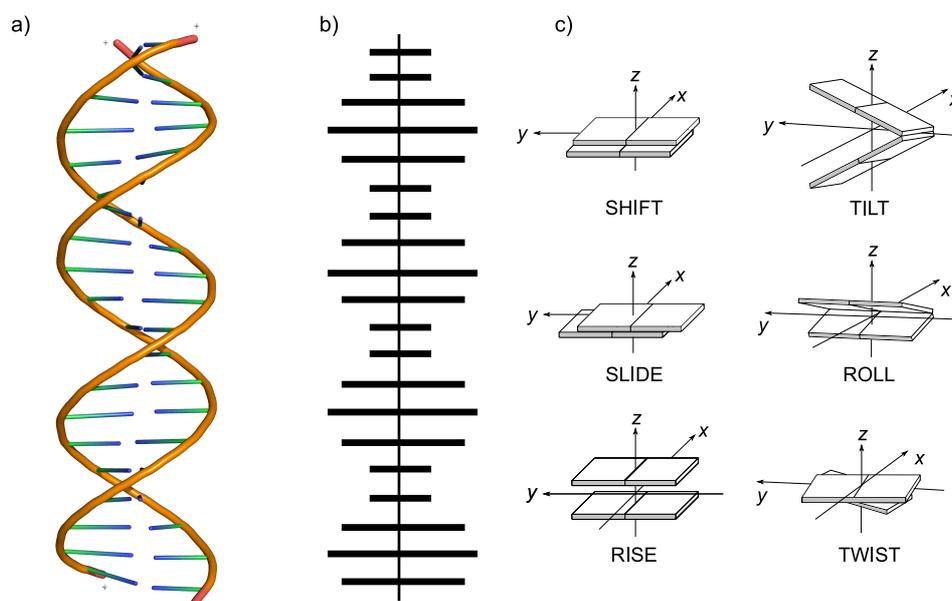
FIGURE 4.10: Coarse graining of DNA. a) Sticks model of an atomistic structure. Each base pair can be coarse grained into a rectangular block, which are assembled into an approximately ladder-like structure shown in b). The relative position of subsequent rectangular blocks is uniquely defined by six base pair step parameters as shown in c).

The base pair step model is ideally suited for extracting the required rigidities at the base pair level [174, 204]. In this model the base pairs are treated as rigid rectangular blocks assembled in a ladder-like manner (Fig. 4.10 a,b). The relative position of two consequent base pairs is defined by six parameters: twist, roll, tilt, shift, slide and rise (Fig. 4.10 c). The definition of these parameters was suggested at a workshop in 1988 [205] and has later been refined and standardised in the scientific community [206]. These standards are implemented in several DNA structure analysis packages; in this thesis we use 3DNA [207].

In 3DNA each base pair is fitted with a plane and assigned a reference frame $[x, y, z]$ so that x lies in the plane of the base and points into the major groove, y also lies in the base-pair-plane and points toward the sequence strand and z points along the helical axis in the 5' to 3' direction of the sequence strand [1]. The angles and distances between consecutive base pairs are measured and translated into six base pair parameters (see Fig. 4.10). Three translations (shift, slide, rise) and three rotations (tilt, roll,

---

[1]The sequence strand is chosen arbitrarily at the start of the calculation. The sequence is always read from the 5' end towards the 3' end. Because the strands are anti-parallel this always results in a right handed coordinate system in the reference frame.

twist) uniquely define the relative position of the base pairs. The parameters are closely connected to twisting and bending rigidities of the elastic rod. The twist parameter is measured in degrees and directly equals the twist angle from Eq. (4.1). Bending of the DNA is anisotropic and is therefore described by two parameters. Bending into the major or minor groove, called the roll, is energetically cheaper than bending perpendicular to the grooves (tilt). However DNA forms a helix with a pitch much smaller ($\approx 3.5\,\mathrm{nm}$) than the bending persistence length of the DNA ($\approx 50\,\mathrm{nm}$). The bending persistence length is thus obtained from a combination of the roll and the tilt parameters.

The relationship between tilt, roll and bend can be derived from the decomposition of the bend angle

$$\mathrm{bend}^2 = \mathrm{roll}^2 + \mathrm{tilt}^2 \tag{4.39}$$

This decomposition is valid for small bend angles, for a proof see [182]. Assuming that tilt and roll are uncorrelated the equipartition theorem states that

$$\frac{1}{2}A_t \left\langle \mathrm{tilt}^2 \right\rangle = \frac{1}{2}k_B T, \tag{4.40}$$

where A_t is the tilting force constant, and similarly for roll and bend, the distinction being that bend has two degrees of freedom. Inserting the equipartition theorem into (4.39) we find that the bending rigidity is then obtained as a harmonic mean of the tilting and rolling rigidities

$$\frac{2}{A_b} = \frac{1}{A_r} + \frac{1}{A_t}. \tag{4.41}$$

We measured the six parameters for each frame of the simulation and then extract local twisting and bending rigidities using a method commonly employed to characterise DNA elastic properties, e.g. [176, 182, 197]. Let us demonstrate the method on three parameters. We denote β_t the tilt, β_r the roll and θ the twist angle. These parameters fluctuate around their means, β_{t0}, β_{r0}, θ_0 respectively and L_0 denotes the mean length of the fragment. The elastic energy is then

$$\frac{E}{k_B T} = \frac{1}{2L_0} \Big[A_r \left(\beta_r - \beta_{r0} \right)^2 + A_t (\beta_t - \beta_{t0})^2 + C(\theta - \theta_0)^2 + A_{tr}(\beta_r - \beta_{r0})(\beta_t - \beta t0)$$
$$+ D_{tt}(\beta_t - \beta_{t0})(\theta - \theta_0) + D_{rt}(\beta_r - \beta_{r0})(\theta - \theta_0) \Big] \tag{4.42}$$

where A_{tr} is the coupling between tilt and roll, D_{tt} coupling between tilt and twist and D_{rt} coupling between roll and twist. The energy can be written in a matrix form

$$\frac{E}{k_B T} = \frac{1}{2}\mathbf{x}^{\mathrm{T}} F \mathbf{x} \tag{4.43}$$

where the deformation vector $\mathbf{x} = \left[(\beta_r - \beta_{r0}), (\beta_t - \beta_{t0}), (\theta - \theta_0)\right]$ and F is the so called stiffness matrix. The probability of the deformation is proportional to the Boltzmann factor

$$P \propto \exp\left(-\frac{1}{2}\mathbf{x}^{\mathrm{T}} F \mathbf{x}\right) \tag{4.44}$$

It can be shown [208] that the elements of the inverse F are equal to the correlations of the elements x

$$\langle x_i x_j \rangle = \left(F^{-1}\right)_{ij}. \tag{4.45}$$

We computed the correlations of all six parameters and constructed the matrix F^{-1} for each of the binding states. The diagonal elements corresponding to twist, roll and tilt were extracted and are shown in Fig. 4.11. The bending rigidities were calculated using Eq. (4.41). All the rigidities are substantially increased in the region of the ligand binding. This increase is much larger than the sequence-dependent variations in the rigidities, validating our assumption of negligibility of sequence dependent effects.

Because these rigidities are extracted from the mean fluctuations we would expect the effective rigidities to decrease slightly in the unbound region, based on the results from the last section. However, we observe a very slight increase. When averaged over the unbound region of the 1:1 complex we observe a 5% increase compared to the same region of the free oligomer, in both the twisting and bending rigidities. This correlation of amplitudes is expected from a positively cooperative system, but in this case this effect must be coming from a different mechanism than the one computed in this chapter.

We evaluate the enhancement factors α and γ, by comparing the rigidity of the free rod and the bound region (Fig. 4.11). We take the average value of these parameters obtained from bp steps 1-3 and 7-9 to find

$$\alpha = 1.3$$
$$\gamma = 1.6. \tag{4.46}$$

The two Hoechst ligands bind to a four base pair sequence with a two base pair gap in the middle, the DNA length is 12 base pairs. We extrapolate the geometrical parameters
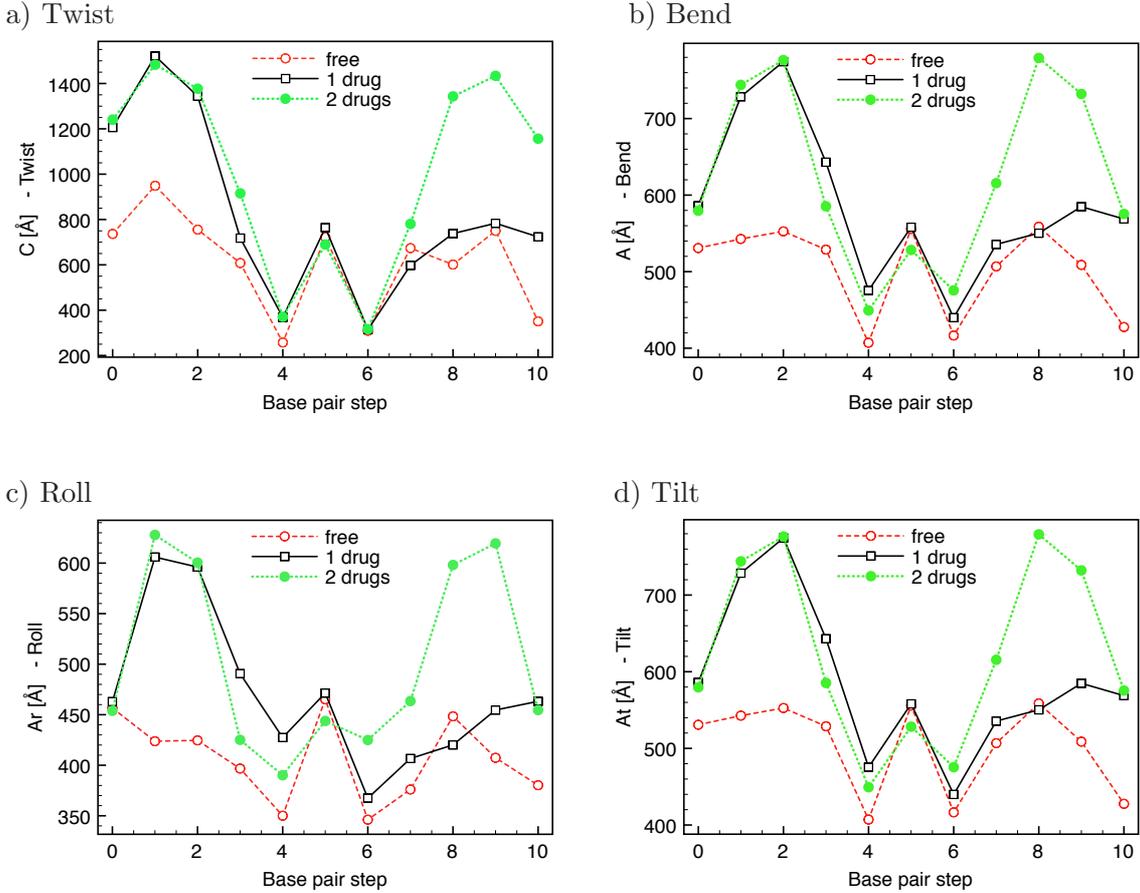
FIGURE 4.11: Twist, bend, roll and tilt rigidities obtained from simulations of Harris el al. [101] plotted for the consecutive base pair steps in three binding states: free DNA (○), 1:1 complex (□) and 2:1 complex (●). The first Hoechst binds between the 2nd and 4th base pairs (if numbering starts at 1), i.e. here the 1st - 3rd base pair step are completely filled by the ligand and the 0th and 4th step are partly filled. The second binding directly affects the 7th - 9th bp step and partly the 6th and 10th bp step.

from this binding pattern to arrive at $a = 1/3$, $d = 1/6$. The number of delocalised modes contributing to allostery is likely to be somewhere between 12 (the number of base pairs in the oligomer) and 70 (the number of atoms on the backbone). We evaluate the free energy for both cutoffs to obtain a plausible range of dynamic contribution to the allosteric effect.

Inserting these values into the formulae for the bending and twisting allosteric free energy we find that $\Delta\Delta G_{twist} = 0.1\,k_BT$ if 12 modes are included and $\Delta\Delta G_{twist} = 0.4\,k_BT$ if 70 modes are taken into account. Bending contributes between $0.02\,k_BT$ and $0.1\,k_BT$

which yields

$$\Delta\Delta G = 0.1\, k_B T - 0.5\, k_B T. \tag{4.47}$$

In other words, the local change of bending and twisting rigidity gives rise to a very weak anti-cooperativity.

Hoechst 33258 was found in experiments and computational studies to bind cooperatively with the allosteric free energy of at least $-6\, k_B T$. The calculations performed in this chapter reveal that the elastic rod model is insufficient in describing the observed positive cooperativity and that we need to go beyond this model and include some more subtle structural fluctuations of DNA in order to capture the desired effect. At the same time the evaluation of the allosteric free energy for the example system is encouraging. The small magnitude of the positive allosteric free energy found suggests that a different dynamic mechanism could be overriding the negative cooperativity arising from modified bending and twisting modes.

4.3 Conclusions

In this chapter we presented a coarse-grained model of dynamic allostery in a rod-like system, with a major application to DNA. The elastic rod is a very different medium for the propagation of the allosteric signal than the previously studied proteins [38, 41, 42, 126]. It is characterised by a large number of spatially delocalised modes that can all become active in the dynamic allostery. The coarse-grained model was therefore chosen so to include all thermally excited, global modes of an inextensible elastic rod. The second important distinction from the previous models was a spatial localization of the ligand of a finite size. The combination of these two factors meant that the binding of the ligand was reflected in a substantially modified structure of the normal modes.

Thermal fluctuations of a typical protein, on the other hand, are known to be composed of relatively small number of slow, global modes separated in the frequency spectrum from a large number of localised fast modes [209]. The coarse-grained models of dynamic allostery are therefore constructed to capture mainly a few dominant slow modes, believed to be crucial for allostery [19, 120]. Binding of ligands in these models leads to nearly or completely unmodified mode structure and tends to be positive cooperative due to the logarithmic dependence of the allosteric free energy on the rigidity (Sec. 2.5,

references [38, 41, 42, 126]). The first binding reduces the amplitude of the fluctuations in the bound but also in the unbound region. The identical change to the rigidity constant representing the second binding is thus entropically cheaper [38]. On the contrary, the binding to the elastic rod is accompanied by an anti-correlated change in the fluctuation amplitude in the occupied and free regions, leading to negative cooperativity.

The allosteric free energy resulting from a perturbation of twisting and bending motions was found to be always positive, signifying an anti-cooperative allosteric effect. In other words, the local perturbation of the twisting or bending rigidity in one segment of the rod increases the free energy penalty for the same perturbation in another segment of the rod. The reason for this anti-cooperative behaviour is that the amplitude of the fluctuations in the unbound sections of the rod is compensating for the effect of the ligand binding on the amplitude of fluctuations in the bound region. That is, if the ligand binding has a local stiffening effect, the amplitude of fluctuations in the ligand occupied region is reduced but the amplitude in the unbound region becomes larger than that of the free rod. It then becomes entropically more costly to increase the rigidity in the unoccupied regions, as illustrated in Fig. 4.9. If, conversely, the ligand binding has a softening effect, the entropy gain upon the second binding is smaller than upon the first, due to the reduced amplitude of fluctuations in the regions surrounding the occupied section of the rod. The allosteric free energy is therefore always positive and the sequential binding anti-cooperative.

As mentioned previously, this is a consequence of the coupling of the eigenmodes from the soft and the stiff region of the molecule. The degree of coupling is proportional to the number of modes close in the frequency spectrum. We have observed a similar effect already in the case of a homodimer where the lowest level of mode coupling was introduced. The uniform elastic properties of the DNA mean that a large number of similar modes is thermally excited and the coupling and the consequent anti-cooperative effect is very strong.

Twisting and bending contributions act through the same qualitative mechanism, but were evaluated separately to allow a quantitative comparison. The character of the dependencies on the free parameters (length of the ligand, separation of the ligands and the rigidity enhancement factor) is therefore the same. If the bending and twisting rigidities are modified in the same regions by the same amount, i.e. $\alpha = \gamma$ and $\alpha \geq 1$ then bending brings a larger contribution to the allosteric free energy. For $\alpha = \gamma < 1$ twisting becomes slightly more important than bending.

4.3.1 Relation to Biological Examples

The negative cooperativity is initially somewhat surprising because to our knowledge only positively cooperative binding has been observed experimentally [20]. This however does not imply that negative cooperativity never occurs, furthermore we have just demonstrated that the effect must be present in the nature to a certain degree. It does, however, indicate that the elastic rod model of DNA, while useful in describing the statistical mechanics of the overall configurations of the polymer, may be inadequate for the more subtle physics of binding correlations. Let us suggest a few possible reasons for the absence of experimental evidence.

First of all, we note that allostery in DNA is very loosely defined and the amount of literature reporting positive cooperativity may thus be slightly misleading. In fact only very few studies have robustly demonstrated positive cooperativity propagated through DNA. We consider DNA allostery in the conventional sense: two identical ligands bind to their respective binding sites on the DNA and the binding of one influences the DNA's affinity for the other. We are interested only in the part of the signal propagated through the DNA. Additionally we assume that the binding is not point-like (thus excluding intercalators) and that it locally modifies DNA bending and/or twisting rigidity over several base pairs. We only consider the effect arising from perturbed thermal fluctuations, which may in principle oppose the effect arising from structural changes. It is thus possible that the anti-cooperative effect studied here has not been observed yet. Let us therefore summarise the available literature reporting on allostery in DNA.

Allostery is most commonly discussed in connection with the assembly of large transcription complexes and enhancers [20, 160, 161]. These are large macromolecular machines, generally of unknown structure, that promote and assist gene transcription. As discussed at the beginning of this chapter there are two mechanisms by which proteins attached to DNA interact. The assembly of these transcription complexes generally belongs to the group of directly interacting proteins. DNA acts more as a scaffolding and its role in the signal transmission is unclear [210, 211].

Another phenomenon often involving allostery is non-specific binding of proteins to DNA. There are several types of such cooperative non-specific interactions. Single strand binding proteins (SSBPs) bind to single stranded regions of DNA and prevent the formation of the duplex [156]. These clearly fall outside of the interest of this chapter. Secondly, a number of proteins, such as lac [156] or λ cl repressor [212], can bind in

a specific as well as non-specific regime. The non-specific regime is often cooperative, resulting in DNA coated with these proteins [200, 210]. However, the experimental results describing the whole binding process are however difficult or impossible to obtain [64, 165, 212]. Furthermore, Diamant and Andelman [165] point out that the coating process may proceed in two phases. They suggest that the first few binding events occur with negative cooperativity but as the overall persistence length of the DNA increases substantially (approximately by a factor of 3) the subsequent protein bindings become cooperative. The initial binding phase however may lie beyond the resolution limit of the experimental techniques

The last example of allostery includes systems where the positive cooperativity in the sense addressed in this thesis occurs. Only a few documented examples of cooperative, specific binding of proteins to DNA exist [210, 213, 214], but better known and more detailed studies have been performed on small minor groove binding ligands such as Hoechst 33258 [101, 155], netropsin [170], distamycin [157], DB183 [171] or small polyamides [159]. We have focused on the example of Hoechst 33258 with the aim to find the mechanism of the positive cooperativity. The elastic rod model proved inadequate for describing the allosteric mechanism and highlights the importance of more subtle structural features of DNA. We conclude that in order to capture the cooperative effect we need to move beyond this simple model and take the more detailed structural fluctuations into account.

The analysis of the trajectories obtained from MD simulations of the H33258/DNA system however showed that the H33258 binding is accompanied by an increase in bending and twisting rigidity. We wanted to evaluate the extent of negative cooperativity arising through this effect and so parameterized our model from these simulations. We used the program 3DNA to extract the local twisting and bending rigidities for the three binding states. These were then inserted into the model and the total allosteric free energy was found to be weakly positive; depending on the number of modes the allosteric free energy was found to equal between $\Delta\Delta G_{twist+bend} = 0.1\,k_BT$ and $\Delta\Delta G_{twist+bend} = 0.5\,k_BT$.

The value between $0.1\,k_BT$ and $0.5\,k_BT$ is relatively small and could be overridden by a different mechanism. There are several possible candidates for the origins of positive dynamic cooperativity. Harris and co-workers [101] for example suggest that the minor groove breathing, disturbed by the binding, may propagate the signal. This hypothesis is supported by the observation that the majority of well-documented cases of positive cooperativity on DNA involve minor groove binding substrates [159, 170, 171]. Other

imaginable scenarios include coupling between elastic modes, such as twist-stretch or twist-bend or anharmonic motions of DNA.

Due to the large number of experimental systems involving the minor groove and due to the availability of the data on the Hoechst/DNA system we investigate the possible role of the minor groove breathing in the dynamic allosteric effects in DNA. The extension of the current elastic rod model to include the minor groove breathing is presented in the following chapter.

Chapter 5

Dynamic Allostery Propagated by DNA Groove Breathing

In the previous chapter we have shown that local perturbations of twisting and bending rigidities can be propagated along the elastic rod and contribute to allosteric effects. However we have found that the only resulting allosteric effect is negative, i.e. if two identical ligands associate with the rod, binding of the first inhibits binding of the second. There are many examples of positively cooperative binding to DNA in nature and although the role of dynamics in the cooperativity has not been assessed in the majority of the cases, it seems unlikely that the dynamics would always act against the cooperativity. Further more, in the example of Hoechst 33258 binding to an oligomer the origin of the positive cooperativity has been shown to be dynamic [101].

As explained at the end of the previous chapter, the experimental evidence for cooperative binding of ligands to DNA is relatively limited. The most detailed results involve cooperative binding of two small, minor groove binding ligands to a short DNA oligomer. In this chapter we first shortly discuss possible mechanisms for cooperative binding to DNA in general and then focus on the example of minor groove binding ligands. We follow the direction laid out by Harris and co-authors [101] and include minor groove breathing into the elastic rod model studied in the previous chapter. The minor groove breathing model is found to give rise to positive as well as negative cooperativity. Both regimes are discussed but more attention is given to the positively cooperative case. In the last sections the model is parameterised from the simulations of Hoechst/DNA system performed by Harris *et al.*

5.1 Possible Mechanisms of Positive Dynamic Allostery in DNA

Several mechanisms can be thought of that could potentially result in positive cooperativity without the requirement of a conformational change. Let us first consider the most obvious: the stretching motion of the elastic rod. Although the stretching modulus of DNA is substantially larger than bending and twisting and is normally not considered in the absence of tension [172, 180, 215], it could be locally modified by the binding of a substrate and this potentially result in an allosteric effect. If however, we inspect the form of the stretching Hamiltonian we find that it is mathematically equivalent to the twisting Hamiltonian Eq. (4.1). Namely, the stretching energy term is

$$E_s = \frac{1}{2} \int_0^L B(s) \left(\frac{\partial u}{\partial s} \right)^2 \mathrm{d}s, \tag{5.1}$$

where B is the stretching rigidity and $u(s)$ is the displacement along the arc length resulting from compression or extension of the rod. This means that qualitatively stretching can only result in the same allosteric effect as twisting, that is negative cooperativity.

Another potential origin of the positive cooperativity is the coupling between twist, bend and stretch. The coupling between twist and bend was disregarded in the previous chapter due to symmetry. Namely, the lowest order coupling term has a general form

$$E_{btw} = \frac{1}{2} \int_0^L D_{btw}(s) \left(\frac{\partial \theta}{\partial s} \right) \left| \left(\frac{\partial^2 \mathbf{r}}{\partial s^2} \right) \right| \mathrm{d}s, \tag{5.2}$$

where $\mathbf{r}(s)$ is the position vector and $\theta(s)$ is the twist angle. For a circular, isotropic beam twisting term in the equation is invariant under rotation by 180 degrees through the axis defined at any point by the tangent vector $\hat{\mathbf{t}} = \frac{\partial \mathbf{r}}{\partial s}$, but the curvature changes sign. Since the energy of any rotated segment remains the same the bend-twist coupling is in this case not allowed. In real DNA this is not the case due to the broken symmetry leading to different minor and major grooves [184]. Similarly, if stretching is included the terms describing coupling between stretching and twisting are generally non-zero [216].

The influence of the substrate binding on the coupling terms is however not intuitive. In order to study the consequences of ligand binding, we determined the coupling constants
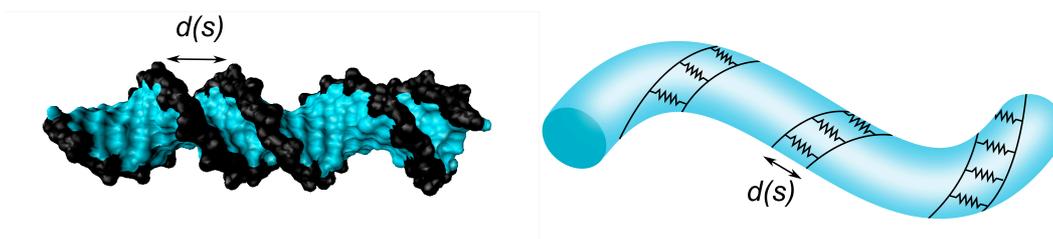
FIGURE 5.1: Left: Surface model of a short piece of DNA. The edges of the grooves are highlighted in black, the minor groove width along the arc length s is denoted d. The arrow connects the i-th phosphorus on one strand and $(i+3)$-rd phosphorus on the opposite strand. Right: The coarse grained model of the DNA as an elastic rod that can bend, twist and includes the groove breathing as a translational motion in a harmonic potential.

from the MD simulations of the Hoechst 33258/DNA system [101] in three binding states, and found that the change of these terms is rather unpredictable (data not shown). We therefore resolve to look for a different source of the positive cooperativity.

5.2 Model of DNA Groove Breathing

As demonstrated by Harris *et al.* ligand binding can disturb the natural "breathing" of the minor groove [101]. Similarly, binding that would interfere with the major groove is expected to perturb the major groove breathing. We therefore include the breathing of the grooves explicitly by extending the model to include such an extra "intrinsic" degree of freedom. We assume that breathing is equivalent to a translational motion of the groove edges as pictured in Fig. 5.1, and that a set of coordinates can be found that decouples this breathing from twisting and bending.

The breathing of the grooves can be viewed as a thermal motion of two chains of atoms with a potential between them. Assuming the atoms in the chain can be represented as point-like beads that are linked by springs and the potential between the chains is harmonic, the groove can be modeled as a beads-and-springs network sketched in Fig. 5.4. The binding of a ligand into a DNA groove is modeled as change of the spring constants in the region of binding.

In this chapter we investigate whether the local perturbation caused by the binding can be propagated along the groove and lead to an allosteric effect. Before we start let us assess the possible degree of cooperativity by studying simple limiting cases of the
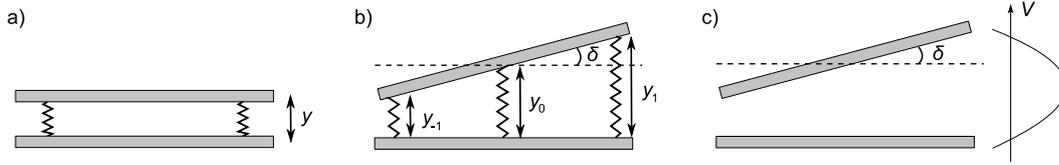
FIGURE 5.2: Toy models of two rigid helices (pictured as grey rods) with harmonic potential between them. a) The helices are held together with a couple of springs and can only move parallel to each other, b) the helices are allowed to tilt, but are connected by three springs in order to display cooperative effect, c) helices that are allowed to tilt move in a continuous potential.

network. We assume for a moment that the groove edges are infinitely stiff and thus the system can be described as two rigid helices, one in a harmonic potential of the other. The two helices thermally fluctuate around their spontaneous parallel configuration at a distance d_0 as shown in Fig. 5.2. The ligand binds between the two helices without modifying their minimal configuration but acts on the spring constants, locally changing them by a factor β. We consider three models of the helices, each of which is found to give rise to positive cooperativity. The values of the allosteric free energy depend on the model. Although the models are trivial they illustrate the rise of the allosteric effects in systems such as the DNA groove.

We consider only one-dimensional helices; translational motion of helices in three dimensions can be written as superposition of motions in three principal directions and the total allosteric free energy is thus three times the allosteric free energy of one dimensional helices.

The simplest model of two rigid helices with two drugs binding consecutively between them is illustrated in Fig. 5.2 c. The helices are connected by two springs and can only move parallel to each other. Each binding modifies one spring constant by a factor β. The derivation of the difference in the free energy change associated with individual spring stiffening is straightforward. Firstly, the energy associated with a deviation y from the relaxed position is

$$E = 2\frac{1}{2}ky^2, \tag{5.3}$$

where k is the force constant of the springs of the unmodified system. Evaluating the free energy of each binding step, first modifying one spring constant by the factor β and
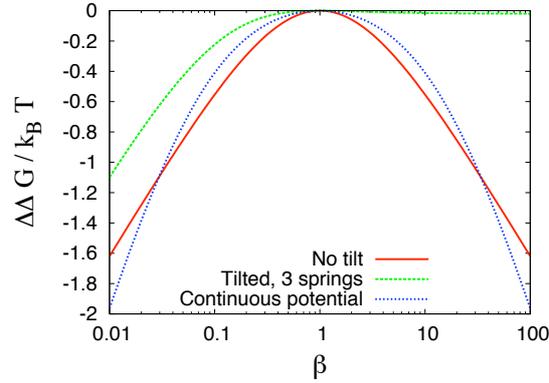
FIGURE 5.3: Allosteric free energy as a function of the enhancement parameter β for three simple rigid helices models (Fig 5.2). For $1/35 < \beta < 35$ the model Fig 5.2 a is the most cooperative, however for $\beta > 35$ or $\beta < 1/35$ the model with continuous potential (Fig 5.2 c) becomes the most cooperative. The model with tilting helices connected by three springs is significantly less cooperative.

then the second spring constant by the same factor we arrive at the allosteric free energy

$$\Delta\Delta G = \frac{1}{2}k_B T \ln \frac{4\beta}{(1 + \beta)^2} \tag{5.4}$$

This term is always negative or equal to zero (for $\beta = 1$), corresponding to positive or no cooperativity. $\Delta\Delta G$ tends to $-\infty$ for large β as is shown in Figure 5.3.

We now extend the model to include tilt between the helices as shown in Fig. 5.2 b. When tilt is added into the model with two springs the allosteric free energy $\Delta\Delta G$ equals zero for all values of parameter β, because the springs act independently [123]. In order to obtain $\Delta\Delta G \neq 0$ a third spring needs to be added to the middle of the rods. The additional spring mediates the communication between the two outer springs. We assume that binding only influences the outer springs, again by a factor β. We derive the allosteric free energy in the same way as for the system with no tilt and find that

$$\Delta\Delta G = \frac{1}{2}k_B T \ln \frac{12\beta(2\beta + 1)}{(5\beta + 1)^2} \tag{5.5}$$

The energy is negative or zero for all values of β, signifying positive or no cooperativity. Compared to the model with no tilt the system is significantly less cooperative for the same values of β (Fig. 5.3) and for $\beta > 1$ quickly approaches its asymptotic value of $\frac{1}{2} \ln \frac{24}{25}$.

Next steps would naturally consist of adding extra springs and observing the effect on $\Delta\Delta G$. Let us skip this and directly proceed to the most refined model with a continuous harmonic potential between the helices (Fig. 5.2 c). Each binding stiffens half of the potential by the factor β, so that the energy of the system with one ligand bound is

$$E = \frac{1}{2}\left(\int_0^{1/2} \beta k y(\hat{x})^2 \, \mathrm{d}\hat{x} + \int_{1/2}^1 k y(\hat{x})^2 \, \mathrm{d}\hat{x}\right), \tag{5.6}$$

and the doubly liganded system is characterised by the harmonic coefficient βk along the whole length of the helices. The allosteric free energy is then

$$\Delta\Delta G = k_B T \ln \frac{16\beta}{\beta^2 + 14\beta + 1}. \tag{5.7}$$

This free energy tends to $-\infty$ faster than the most simple model with no tilt, but is less cooperative for $1/35 \lesssim \beta \lesssim 35$, see Fig. 5.3.

Most importantly these results indicate that this mechanism can be the source of positive cooperativity and mark out the limiting values of expected allosteric free energies.

5.2.1 Method

Let us generalise the "stiff helix" family of models, discussed above, to a more realistic model of DNA groove. A DNA groove is approximately composed of two chains of atoms forming the groove edges with an effective potential between them, similar to the potential between the stiff helices in the previous section. The groove can be represented by a bead and spring model shown in Fig. 5.4, which, in contrast to the stiff helices, captures the thermal fluctuations of atoms within the edges.

The interaction within the edges is modeled by harmonic springs of identical force constant $\tilde{k}$. The potential between the edges is coarse-grained into harmonic springs between the opposite beads on the backbones. All springs simulating the interaction of the two edges have a force constant $\tilde{\lambda}$ (Fig. 5.4). We define a set of coordinates so that the x-axis points along the main axis of the groove, y-axis is parallel to the coupling springs in the relaxed position. We can consider the network to exist only in the plane since the motions in the three principal directions are decoupled, as we show below. Any extension of a spring in this plane can be decomposed into its x and y components. The energy
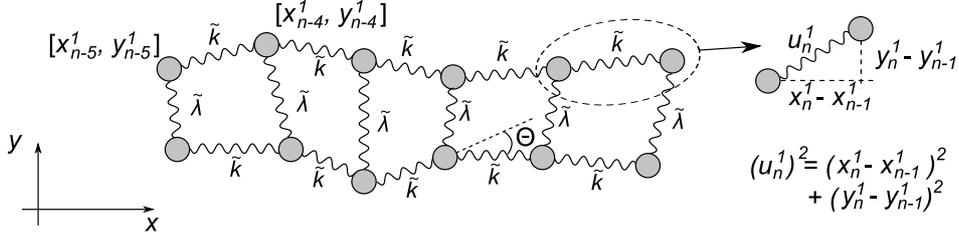
FIGURE 5.4: Bead and spring network model of the DNA groove.

of the bead and springs model is then

$$E = \sum_{i=1}^{2} \sum_{n=1}^{N} \frac{1}{2} \tilde{k} \left(\left(x_n^i - x_{n-1}^i \right)^2 + \left(y_n^i - y_{n-1}^i \right)^2 \right)$$
$$+ \frac{1}{2} \tilde{\lambda} \left(\left(x_n^1 - x_n^2 \right)^2 + \left(y_n^1 - y_n^2 \right)^2 \right). \tag{5.8}$$

The x and y coordinates are decoupled which means that the transverse and longitudinal motions are independent and their respective equations of motion identical. Without a loss of generality we concentrate on the transverse motion only. The same network in three dimensions results in three independent but mathematically identical equations.

Let us now simplify the model by assuming the atoms are so close to each other that each chain can be treated as continuously flexible and the finite number of springs connecting the chains can be replaced by a continuous harmonic potential. Denoting the transverse displacement of the individual chains as y_1 and y_2, the continuous analog of Eq. (5.8) is then

$$E_{trans} = \int_0^L \left(\frac{1}{2} k \left(y_1'^2 + y_2'^2 \right) + \frac{1}{2} \lambda (y_1 - y_2)^2 \right) ds, \tag{5.9}$$

where k and λ are scaled spring constants $\tilde{k}$ and $\tilde{\lambda}$, which fulfill $k = \tilde{k}/\Delta x$ and $\lambda = \tilde{\lambda} \, \Delta y$, where Δx and Δy are the equilibrium distances between the neighbouring beads on the backbones and on the opposite backbones respectively.

This equation can be further simplified by introducing two new variables: the inter-backbone distance $f = y_1 - y_2$ and the average backbone displacement $g = y_1 + y_2$. Equation (5.9) then becomes

$$E_{trans} = \int_0^L \left(\frac{1}{4} k \left(f'^2 + g'^2 \right) + \frac{1}{2} \lambda f^2 \right) ds. \tag{5.10}$$

The normal modes of this Hamiltonian for k = const. and λ = const. are found similarly to the twisting normal modes of the elastic rod (see Appendix C) and yield

$$f = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi s}{L}\right) \qquad g = \frac{b_0}{2} + \sum_{n=1}^{\infty} b_n \cos\left(\frac{n\pi s}{L}\right). \tag{5.11}$$

In this case we have to include the constant terms because of the presence of the f^2 term in the Hamiltonian.

The ratio k/λ has a dimension of a length squared. The meaning of this length can be found by introducing a perturbation to f at the origin and minimizing E_{trans} with respect to f. This requires solving a differential equation

$$k f'' - 2\lambda f = 0 \tag{5.12}$$

with the boundary conditions

$$f|_{x=0} = f_0 \qquad\qquad f|_{x=\infty} = 0 \tag{5.13}$$

The solution is an exponential decay function

$$f = f_0 \exp\left(-\sqrt{\frac{2\lambda}{k}}s\right) \equiv f_0 \exp\left(-\frac{s}{\xi}\right) \tag{5.14}$$

with the decay length $\xi = \sqrt{k/2\lambda}$. Any perturbation to the relaxed width of the system thus decays by a factor of e over the decay length ξ. This length scale becomes important when considering a finite separation of the binding ligands.

We remove the L dependence from the energy expression by introducing the scaled variable $x = s/L$. The number of free parameters is further reduced by introducing a dimensionless ratio ξ^2/L^2. The energy then becomes

$$2\frac{E_{trans}}{L\lambda} = \int_0^1 \left[\frac{\xi(x)^2}{L^2}\left(f'(x)^2 + g'(x)^2\right) + f(x)^2\right]\mathrm{d}x, \tag{5.15}$$

where the decay length ξ is in principle a function of the coordinate x.

We now evaluate the allosteric free energy associated with the sequential binding of two ligands between the chains (into the groove). Assuming that each binding has identical effect on the force constants, the attachment of the ligand results in the constant k
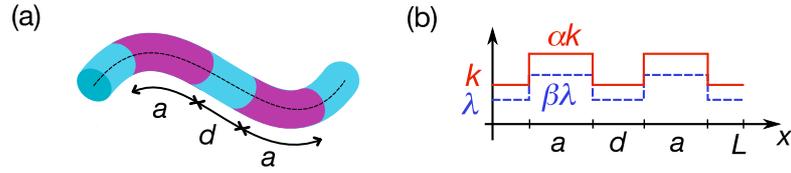
FIGURE 5.5: (a) A schematic illustration of the geometry of doubly bound DNA and (b) the effect of the ligand on the elastic parameters of the groove. The location of the ligands is pictured in purple.

altered by a dimensionless factor α and λ becoming $\beta\lambda$ in the region with the ligand bound. We refer to α as the phonon enhancement or stiffening factor and to β as the groove stiffening factor. For example when the ligand stretches across a half of the groove the energy becomes

$$E_{trans} = \frac{\xi^2}{L^2} \left(\alpha \int_0^{1/2} + \int_{1/2}^1 \right) \left(f'^2 + g'^2 \right) \mathrm{d}x + \left(\beta \int_0^{1/2} + \int_{1/2}^1 \right) f^2 \mathrm{d}x. \tag{5.16}$$

Constants of proportionality that cancel in the expression for $\Delta\Delta G$ have been omitted. A general geometry is implemented by setting the appropriate limits on the integrals. In this chapter, we continue using the geometrical parameters (a for the ligand length, d denotes the ligand spacing and L is the length of the rod) introduced in the last chapter. Figure 5.5 shows schematically the doubly bound DNA and illustrates the definition of the parameters.

The energy is then rewritten in terms of normal modes by inserting Eq. (5.11) into the energy expression. $\Delta\Delta G$ is evaluated in the same manner as for bending and twisting, with caution paid to the truncation of the matrices (see Sec. 4.2.2 and Appendix D).

It is worth noting that in this system modeling the groove-width modes is elastic but not stiff, i.e. the persistence length of the groove width is zero. As we shall see from the example of Hoechst 33258 binding to DNA, this is a reasonable assumption for the DNA groove. We measured the DNA groove width fluctuations of this system in all three binding states. The results are shown in Fig. 5.6 a and strongly suggest that the persistence length of the groove edges is negligible. The binding of the ligand substantially reduces fluctuations in the occupied region but this effect is not propagated along the DNA. This signifies very soft edges, whose stiffness can be neglected.
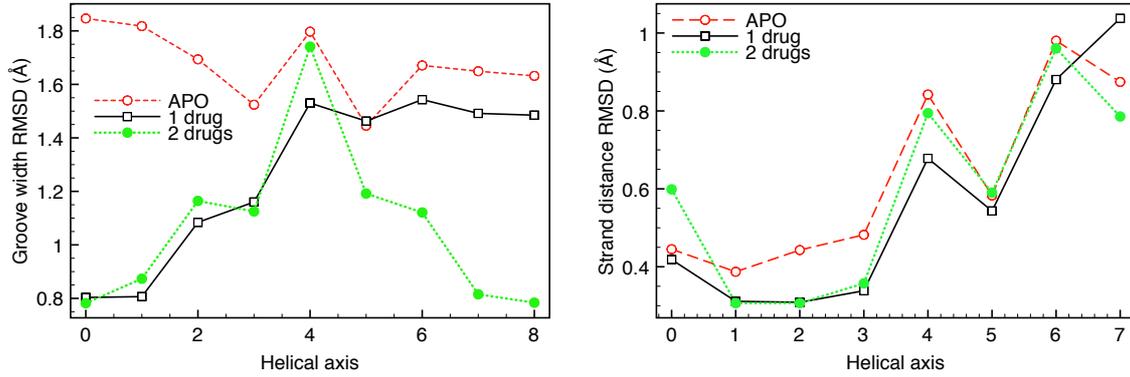
FIGURE 5.6: Fluctuation of the DNA minor groove measured during the association with H33258 in the simulations of Harris *et al.*[101]. The fluctuations were measured between the i-th and $(i+3)$-rd phosphorus located on the opposite edge (Fig. 5.1). Apo (○) corresponds to a free dodecamer, 1 drug signifies one H33258 bound to the left hand side of the dodecamer (□), and 2 drugs denote two H33258 bound with one base pair left unclamped in the middle (●). The groove width fluctuations are reduced only in the region of binding. The strand fluctuations, measured along each groove edge, are not affected, signifying that the factor α of our model is equal to one.

Nevertheless in some cases the stiffness may play an important role and has to be explicitly included in the model. Let us only sketch the calculation in such a case. In the soft elastic beads and springs model the beads constitute hinges through which the bonds are allowed to rotate freely. Introducing the stiffness into the model is equivalent to penalizing this rotation. The angle of rotation (denoted by Θ in Fig. 5.4) can be related to displacements x_n and y_n. For small deviations this results in an additional term in Eq. (5.9) which reads

$$\int_0^L \frac{1}{2} A \left(y_1''^2 + y_2''^2 \right) \mathrm{d}s. \tag{5.17}$$

The eigenmodes of this system can be written as a combination of trigonometric and hyperbolic functions similar to the solution of the bending motions in the elastic rod model of double-stranded DNA that we explored in chapter 4. We can then apply the same procedure that was used for the bending and twisting to obtain the allosteric free energy. Instead of performing the full calculation we estimate the maximal obtainable cooperativity from the rigid helices model with the tilting helices and a continuous potential between them (Fig. 5.2 c). The resulting $\Delta\Delta G$ is plotted in Fig. 5.3 and shows that the the limiting cases of the infinitely stiff edges (in blue) and the edges with zero

persistence length (in red) are very similar. This is true particularly for biologically relevant small and intermediate values of β.

5.2.2 Results

We have derived the method for calculating the allosteric free energy associated with a perturbation of the DNA groove breathing (Eq. (5.15)) under the assumption of soft edges and a harmonic potential across the groove. The allosteric free energy is a function of the geometric parameters a (the ligand length) and d (distance between the ligands) and the elastic parameters α, β and ξ. The decay length ξ is a measure of flexibility of the groove; small ξ signifies more flexibility and faster decay of the allosteric signal. Let us first study dependencies on these parameters in general and compare them to the toy models presented earlier (Fig. 5.2). Subsequently we take a concrete example of an allosteric system in the form of Hoechst 33258 binding to a DNA oligomer and parameterise our model from the MD simulations of this system.

The allosteric free energy derives from the energy expression (Eq. (5.9)) which involves two parts: one associated with stretching of the edges and second associated with the

a)

b)
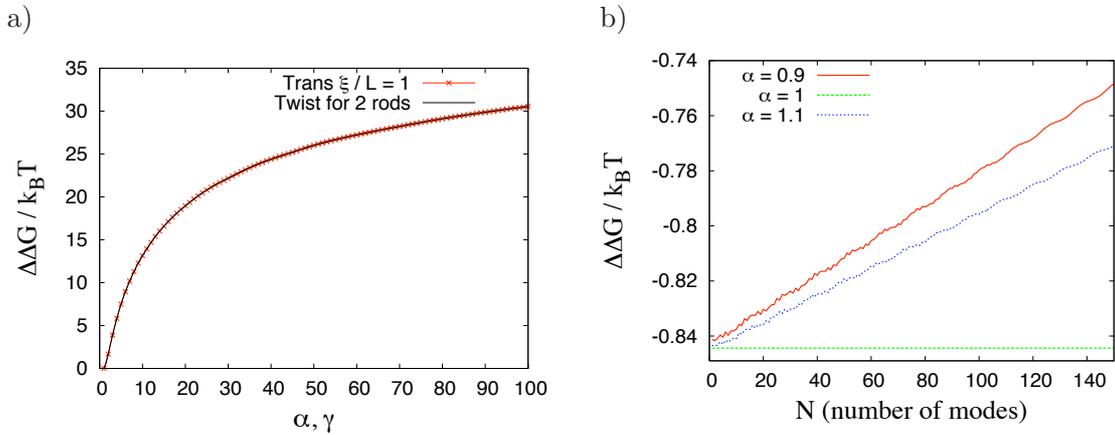
FIGURE 5.7: Allosteric free energy associated with the groove breathing as a function of a) phonon enhancement factor α and b) number of modes. The $\Delta\Delta G_{trans}$ in a) corresponds to simple stretching or twisting of two rods as is demonstrated by plotting the result for twisting alongside the translational free energy. In b) the $\Delta\Delta G_{trans}$ converges to a constant when $\alpha = 1$, but for $\alpha \neq 1$ every mode added to the system decreases the cooperative behaviour. The geometry used is defined by $a = 1/3$, $d = 1/6$ and $\xi/L = 1$. In a) the groove stiffening factor β was set to unity and in b) $\beta = 30$.
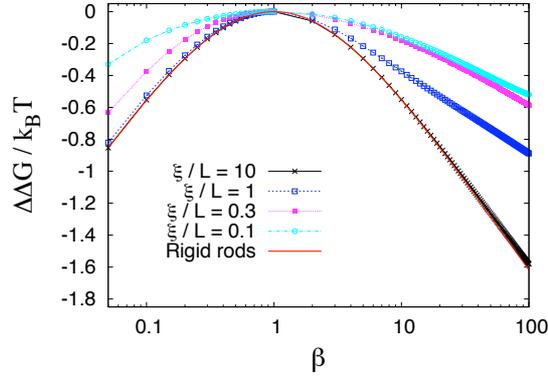
FIGURE 5.8: Allosteric free energy for $\alpha = 1$ as a function of the groove stiffening factor β. The breathing groove model approaches the rigid helices model (in red) for $\xi/L \to \infty$. The groove potential is modified in each of its two half-lengths sequentially (i.e. $a = 1/2$ and $d = 0$).

potential between the edges. The two parts are decoupled in the energy expression and hence the allosteric free energy is in the first approximation a sum of two independent processes. Hence, they can be studied separately without a loss of generality. Let us start by setting $\beta = 1$ or equivalently keeping the potential inside the groove unchanged upon ligand binding. The allosteric free energy in this case is identical to stretching of two rods and perturbing their stretching modulus by the ligand binding. We know that energy of a stretched rod is formally identical to energy of a twisted rod. In Fig. 5.7 we plot twice the twisting allosteric free energy obtained in Chap. 4 together with $\Delta\Delta G_{trans}$ for $\beta = 1$ and observe that the two curves are virtually identical.

The second process is highlighted by setting the phonon enhancement factor $\alpha = 1$ and studying the β dependence. This corresponds to keeping the extensibility of the groove edges constant upon the ligand binding and changing only the potential between the edges. In this case positive cooperativity occurs, as expected from the rigid helices models. The positive cooperativity is predictably larger for longer decay length (large ξ/L). The asymptotic behavior is compared to the rigid helices in Fig. 5.8. For the comparison the geometry has been chosen so that the groove stiffens in each of its two half-lengths sequentially ($a = 1/2$ and $d = 0$). For a large decay length ξ, the allosteric free energy of the elastic groove approaches the energy of the limit case of two parallel rigid helices. No tilt, analogous to the other two rigid helices models, occurs in this case due to the edge persistence length having been set to zero.

a) $\alpha = 1$, $\beta = 100$

b) $\alpha = 100$, $\beta = 1$
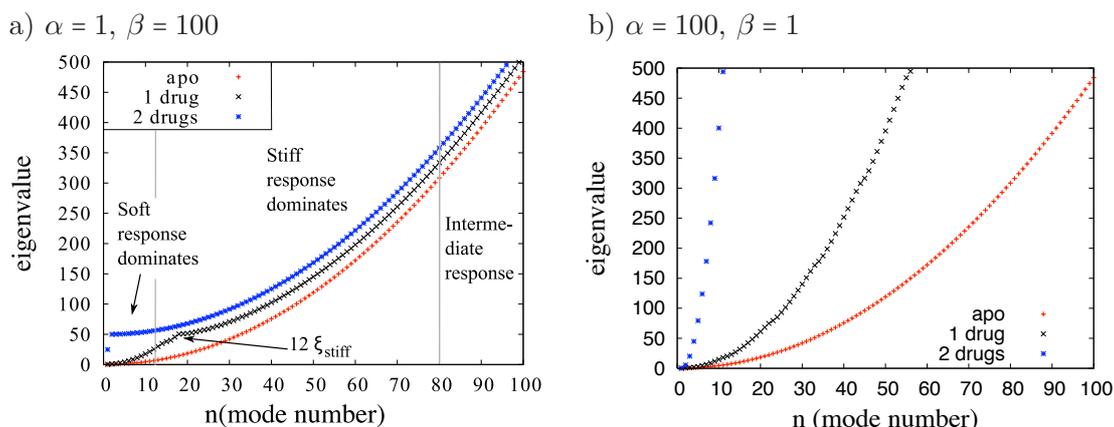
FIGURE 5.9: Eigenvalues of the energy matrices of the three binding states corresponding to homogeneous potential along the whole groove (apo and 2 drugs), and a half of the groove stiffened by parameters α and β (1 drug). a) The parameters are $\alpha = 1$, $\beta = 100$ and $\xi/L = 0.1$. In the half stiffened groove we observe the eigenmode structure changing from one dominated by soft response to one dominated by stiff response and eventually to an intermediate response where neither half dominates. b) $\xi/L = 0.1$, $\alpha = 100$ and $\beta = 1$; the three eigenvalue sequences diverge for increasing n.

When both α and β differ from unity the $\Delta\Delta G$ is a result of the interplay between the two mechanisms. Comparing figures 5.7 and 5.8 we observe that the negative cooperativity arising from the phonon alteration ($\alpha \neq 1$) is much larger than the positive cooperativity from the groove stiffening ($\beta \neq 1$). In total, positive cooperativity is only observed for values of α around unity. As seen in Fig. 5.7 b, if $\alpha \neq 1$ then $\Delta\Delta G_{trans}$ increases linearly with the number of modes included in the system. This is due to the mechanism, discussed in the previous chapter, that causes the same dependence on number of modes in $\Delta\Delta G_{twist}$. When $\alpha = 1$ the allosteric free energy quickly converges to a constant with increasing N.

The dependence of the allosteric free energy on the number of modes gives an insight into the physics of the signal propagation. The physical principles become most apparent when the eigenvalues of individual energy matrices are evaluated as a function of α, β and ξ/L. Example sets of eigenvalues are plotted in Fig. 5.9.

Several observations can be made: for $\alpha = 1$ the edges of apo and 2:1 complex are homogeneous and differ only in the strength of their interaction. Let us for the moment assume that the drug binding stiffens the groove potential, i.e. $\beta > 1$. The increased interaction in the 2:1 complex is manifested by uniformly increased eigenvalues compared to the apo structure. An interesting structure of the eigenvalues arises for the half
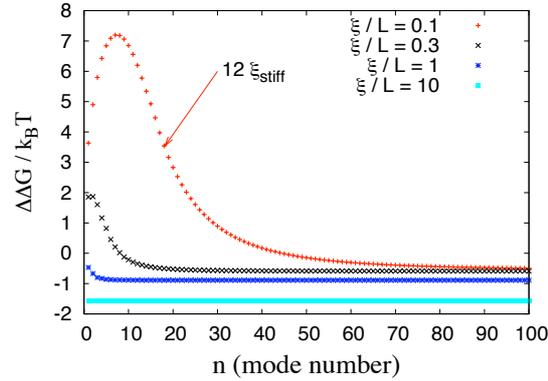
FIGURE 5.10: Allosteric free energy as a function of number of normal modes for $\alpha = 1$ and $\beta = 100$. We see that the allosteric free energy converges to a constant value for a large number of modes.

stiffened system. Three regimes are observed in this case: 1) for low mode numbers (long wavelengths), the soft response dominates. That is, the mode structure is predominantly determined by the mode structure in the softer, unbound half of the system. The influence of the stiffer part increases with the mode number and eventually becomes dominant. The dominance of the stiff response constitutes the second regime. The increasing dominance of the stiff response continues for a few more modes until a sharp kink occurs (Fig. 5.9). We studied the location of the kink and found that it is linked, as we might expect, to the decay length ξ_{stiff} in the stiffer part of the rod.

If the phonons remain unperturbed ($\alpha = 1$) the singly-bound system is composed of two parts differing only by the strength of the interaction between the edges. Any perturbation at the interface of the two regions therefore decays across a different length in each half: in the drug free region (soft part) the decay length $\xi_{soft} = \sqrt{k/2\lambda}$ and in the stiffer part $\xi_{stiff} = \sqrt{k/2\beta\lambda}$. The kink coincides with a mode whose wavelength equals 12 times the stiff decay length ξ_{stiff}. The kink constitutes a turning point from which the dominance of the stiff response decreases until the eigenvalue of the half stiffened system is exactly half of the difference between the apo and the 2:1 eigenvalue. The modes for which this is true do not contribute to the allostery. This results in a convergence of $\Delta\Delta G$ with the number of modes N shown in Fig. 5.10.

The same conclusions apply when the drug binding has a softening effect on the groove potential, i.e. $\beta < 1$. The soft response dominates for the low mode numbers and stiff response dominates for medium mode numbers.
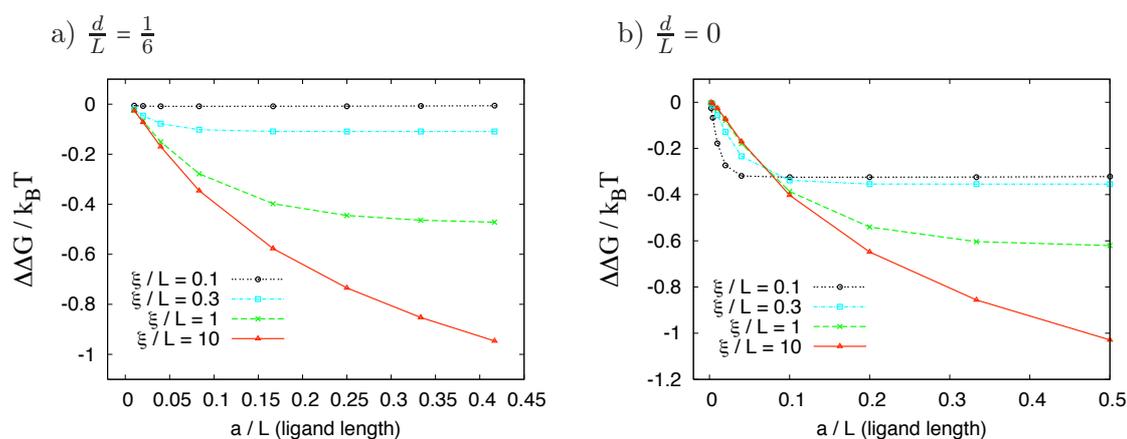
FIGURE 5.11: Allosteric free energy as a function of the normalised ligand length a for the spacing between ligands (a) $d/L = 1/6$ and (b) $d/L = 0$. Other parameters used were $\alpha = 1$ and $\beta = 30$. The cooperativity increases with a but becomes saturated when $a = \xi/2$. For $d = 0$ and small a the system becomes more cooperative for smaller decay length ξ.

When $\alpha \neq 1$ the eigenvalue sequences fan out as shown in Fig. 5.9. The value of ξ/L determines the steepness of the growth of eigenvalues with the mode number. The larger value of ξ/L corresponds to a steeper growth which in turn leads to an amplified allosteric effect. If $\beta \neq 1$ the mode structure with the kink described above is still present but the anti-cooperative effect coming from phonon modification is significantly larger.

Because phonon stiffening ($\alpha \neq 1$) has been studied before, we set $\alpha = 1$ for the rest of this section. This can also be justified as a reasonable assumption for many biological systems. When a ligand binds between two helices, or into a DNA groove the elasticity of the helices or the groove edges is likely to change much less than the interaction across the gap between the helices, or across the groove. It can therefore be neglected or equivalently α assumed equal one.

The geometry of the system is reflected in the values of $\Delta\Delta G$ in a more complicated and yet more realistic way than in simple bending or twisting, due to the additional and non-trivial dependence on the intrinsic decay length ξ. The dependence on the length of the ligand a normalised to the DNA length L is shown in Fig. 5.11. As expected the cooperativity increases with the ligand length a, but becomes saturated when a exceeds approximately half the decay length ξ.
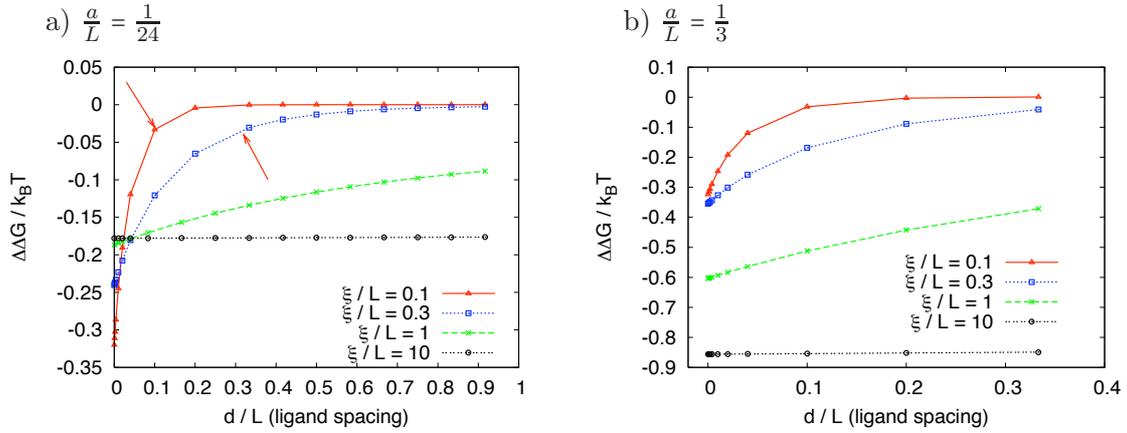
FIGURE 5.12: Allosteric free energy as a function of the normalised distance between ligands d/L for a) the ligand length $a/L = 1/24$ and b) $a/L = 1/3$. Other parameters used were $\alpha = 1$ and $\beta = 30$. For $a/L = 1/24$ and small d/L the groove characterised by a small decay length ξ/L becomes more cooperative. Red arrows in a) mark the decay length ξ/L.

The distance ξ can be understood as a penetration depth of the allosteric signal. The perturbation caused by the first ligand binding is propagated over the distance ξ and any binding within this range then gives rise to allostery. Any change of the mechanical properties beyond this distance has only negligible effect on the allosteric free energy, resulting in the plateau observed in Fig. 5.11 for the low values of ξ/L. For larger values of ξ/L the plateau occurs at distances not included in the figure. The saturation effect of $\Delta\Delta G$ is illustrated in a more systematic manner in Fig. 5.13 a. Here the allosteric free energy is plotted as a function of the dimensionless ratio a/ξ. The allosteric free energy saturates when $a = \xi/2$, i.e. when the length of the bound region in the doubly bound state equals the penetration depth.

In Fig. 5.11 we also notice that for d very small with respect to L, instead of the anticipated increase of cooperativity with increased penetration depth, the reverse is found for small a. It becomes more advantageous for the second, short, ligand to bind if the decay length ξ/L is smaller.

We see the same type of dependence if we plot $\Delta\Delta G$ as a function of the normalised distance between ligands d/L (Fig. 5.12). The decay length ξ/L is marked in the figure and represents a point where the maximal allosteric effect obtained for $d = 0$ is reduced by a factor e. The scaling of the allosteric free energy with the penetration depth ξ/L is shown in Fig. 5.13. It illustrates the identical character of the dependencies, when

scaled properly, but conceals the behavior at small a/L and d/L. Fig. 5.13 shows that if $a \gg \xi$ the ligands are sensed as infinitely long by the allosteric signal and the curves for different ξ/L collapse onto one. Similarly if $a \ll \xi$ the ligands are perceived as point-like perturbations and the resulting curves are identical too (data not shown).
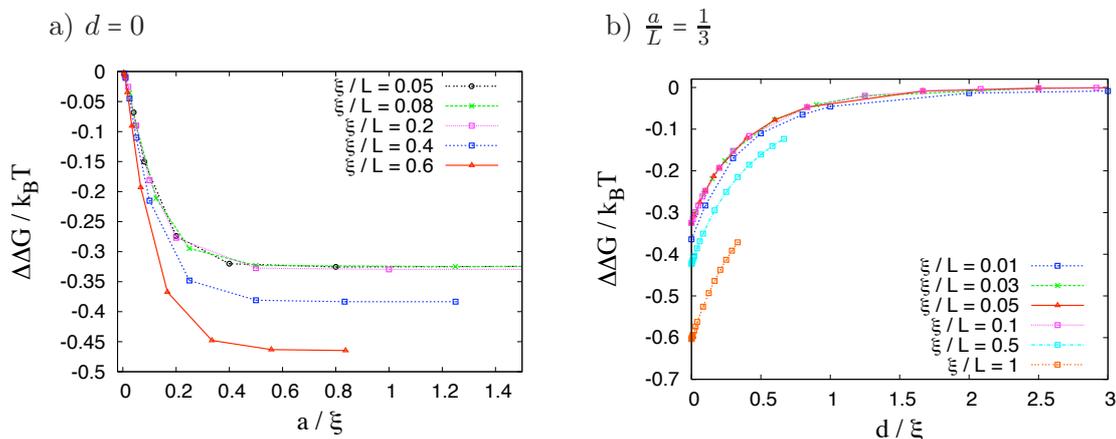
FIGURE 5.13: Allosteric free energy as a function of the dimensionless ratios (a) a/ξ and (b) d/ξ. The plots show that the character of the dependence of the allosteric free energy on the distances is identical for all ξ.

5.2.3 Application to the DNA/Hoechst 33258 System

Harris *et al.* suggest that the minor groove breathing plays a crucial role in the allosteric effect of Hoechst 33258 binding to the DNA oligomer. H33258 fits very tightly into the minor groove and substantially perturbs the breathing of the groove (Fig. 4.1 and 4.2). We used the 11 ns long repeat of their simulations to extract the parameters required to evaluate our model.

The values of geometrical parameters were estimated from measuring the Hoechst occupied region of the groove in the program VMD [217]. We averaged over 10 frames, spaced by 1 ns in the simulations to find the parameters

$$a/L = 0.435 \tag{5.18}$$

$$d/L = 0.13. \tag{5.19}$$

In order to determine the elastic parameters (ξ, α and β) for this system we measured the minor groove width f for each frame of the simulation as the distance between the i-th
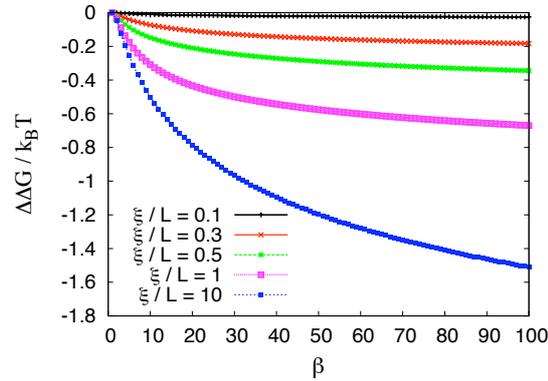
FIGURE 5.14: Translational allosteric free energy for the geometry of the DNA-H33258 system (a = 0.435, d = 0.13). The flexibility of the groove edges is assumed to remain constant during the binding (α = 1).

and $(i+3)$-rd phosphorus on the opposite backbones (Fig. 5.1). The elastic parameters were then extracted via a method involving decomposition of the widths into Fourier modes. The details of method with the fits to the data are presented in Appendix E.

The method rests on fitting of the expected function of the average groove width into the measured values. The best fit resulted in the following set of parameters ξ/L = 0.06, α = 0.5 and β = 6.3, but ξ/L between 0.06 and 1.5 and α = 1 were also found to give reasonable results. To summarise, the method confirmed that stiffening of the groove is the main consequence of the ligand binding and that phonons become either slightly softer or remain unchanged. The fits were too unreliable to determine the exact value of ξ/L, and therefore we resolve to estimate the plausible $\Delta\Delta G$ based on physical arguments.

In order for the two drugs separated by two base pairs to be able to communicate we estimate the decay length to be at least 3-4 bp ($\xi/L \in (1/4, 1/3)$). Since the drug fits into the minor groove very tightly and thus increases the groove potential substantially, we estimate that β may be in the range of 10-30. Fig. 5.14 shows that in such case the allosteric free energy ranges from $-0.05\, k_BT$ to $-0.15\, k_BT$, corresponding to $-0.15\, k_BT$ to $-0.45\, k_BT$ for three degrees of freedom of a real molecule.

These values are relatively insensitive to the number of modes contributing to the allosteric effect. We chose to use the asymptotic values (see Fig. 5.10), which are reached for 8-25 modes depending on ξ/L and β. If only 12 modes are included, corresponding to the shortest wavelength equal to the distance between the base pairs, the range of

plausible free energy is modified only slightly; $\Delta\Delta G$ is found to be between $-0.15\,k_BT$ and $-0.4\,k_BT$ for three degrees of freedom.

The free energy values are still very low compared with the experimental ($\Delta\Delta G = -6.8\,k_BT$) and computational ($\Delta\Delta G = -11.8 \pm 1.3\,k_BT$) results. In addition, in the first part of this chapter we found that bending and twisting act against the cooperativity and contributes an allosteric free energy of $+0.5\,k_BT$, which would effectively cancel this result.

5.3 Fast Modes

There is one more possible source of cooperativity arising from harmonic modes that we consider in this thesis. Hawkins and McLeish (ref. [42] or Sec. 2.5) have shown that in some special cases enslaving of fast modes can amplify the allosteric effect. They focused on a discrete model with the eigenmode structure unperturbed by binding, and concluded that if the depth of the fast mode potential remains unaffected by the binding, the allosteric free energy can be amplified up to M-fold, where M is the number of enslaved modes (Sec. 2.5).

In the case of DNA, the localised, thermally excited modes can be, for example, fluctuations of individual bases. As a proof of principle, we examined the individual base fluctuations of our test case, the DNA/H33258 system. We analysed the 11 ns MD simulation using the program 3DNA introduced in Sec. 4.2.3.4. 3DNA measures the six intra base pair parameters (shift, tilt, slide, roll, rise and twist) introduced earlier but is also able to describe the base pair geometry in more detail. Base pairs are in reality not planar as is assumed in the rigid base pair model. The relatively rigid bases are connected by two or three comparatively weak hydrogen bonds and consequently move with respect to each other. The relative position of the bases within a base pair can be uniquely described by six parameters; three rotational (buckle, propeller twist and opening) and three translational parameters (shear, stretch and stagger). For illustration refer to Fig. 5.15. These six inter base pair parameters together with six intra base pair parameters uniquely define the position of each base in the DNA.

The intra base pair parameters were measured for each frame of the simulation. The RMSDs of the parameters were determined and the dependence on the binding state studied. The results are presented in Fig. 5.16. Shear, stagger, buckle and propeller
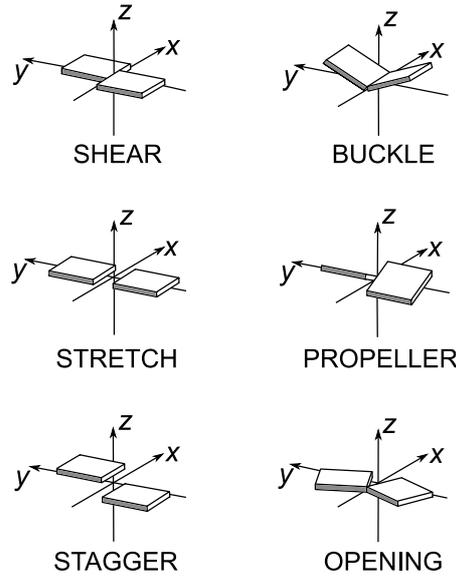
FIGURE 5.15: Pictorial definition of the base pair parameters. Construction of the coordinate frame is identical to the base pair step parameter model explained in Sec. 4.2.3.4.

display reduced amplitude on 1-3 base pairs per binding site. Stretch and opening, on the other hand display a weak loosening in the occupied regions. The terminal base pairs are too unstable to be considered.

It is possible that these localised modes are enslaved by the delocalised modes and amplify allostery. The general model of enslaved fast modes has been introduced in Sec. 2.5. Here we perform the calculation for the continuous elastic medium of DNA and study the possible amplification in this case. We assume that both the depth and the width of the fast mode potential is coupled to the slow mode. As we explained earlier the choice of the coupling function is arbitrary so long as it describes the desired effect: the fast mode potential becomes shallower and flatter with increased amplitude of the slow mode (Fig. 5.17). In chapter 3, when addressing allostery in homodimers, we identified coupling functions that are convenient for dealing with multiple slow modes and we continue to use them here.

The coupling between the slow and fast modes is reflected in new terms in the expression for the energy of the system. The energy becomes a sum

$$E = E_{slow} + E_{fast}, \tag{5.20}$$

a) Shear

b) Stretch

c) Stagger

d) Buckle

e) Propeller
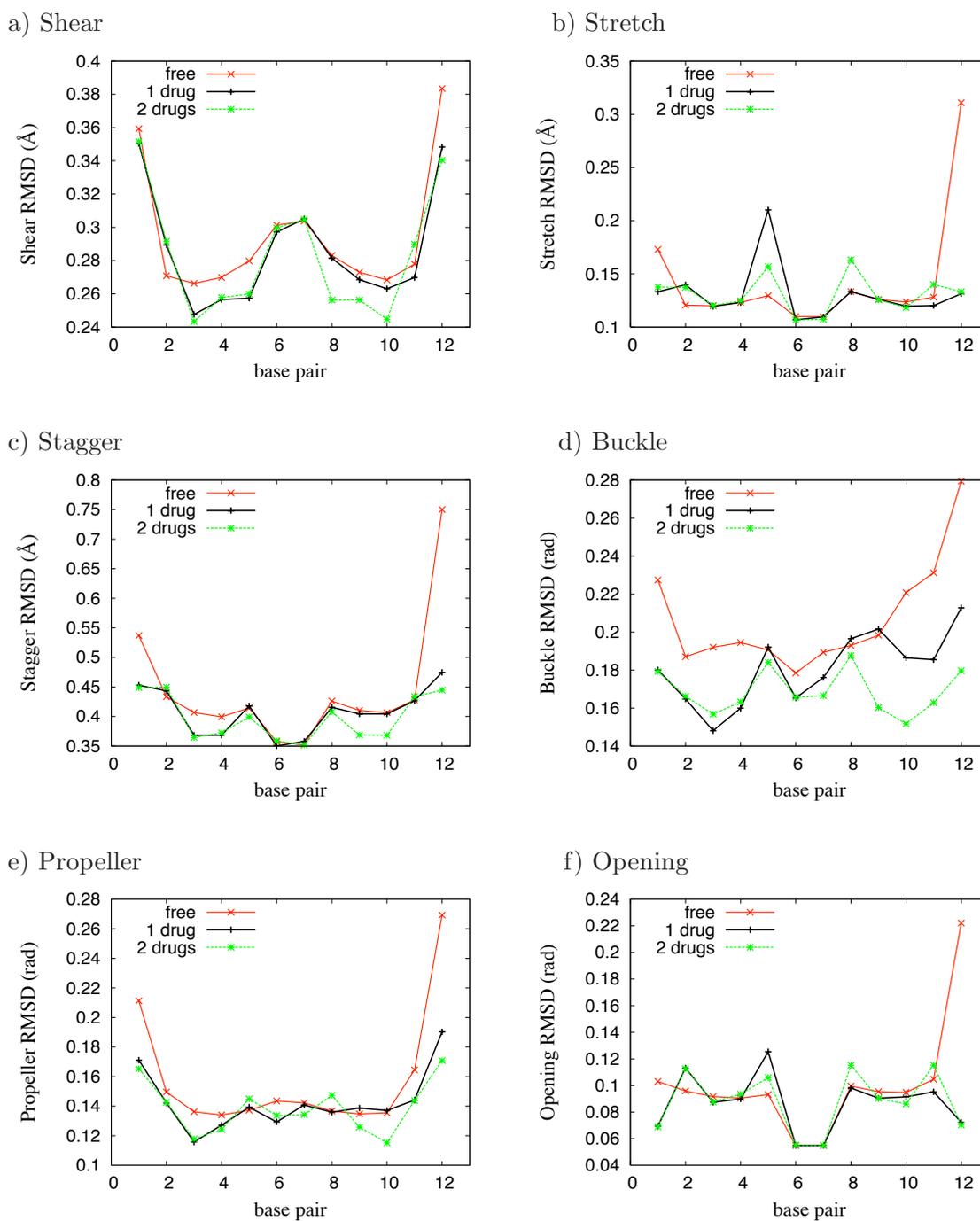
f) Opening

FIGURE 5.16: Root mean square deviation of the six base pair parameters measured for three binding states of the DNA/H33258 complex: free (×), 1:1 (+) and 2:1 (✳) state.
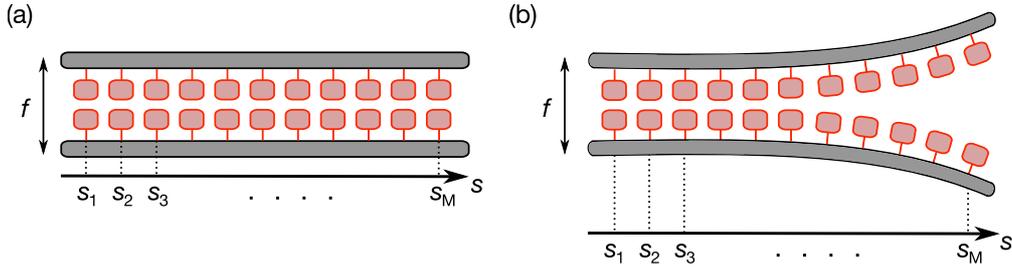
FIGURE 5.17: Schematic illustration of fast, localised modes enslaved to the groove opening. Localised fluctuations of a small structure, here pictured in red, are coupled to the groove width f at a single point along the groove denoted as s_i. (b) As the groove becomes wider, the small structures (in red) have more room for fluctuations, hence their potential becomes shallower and wider.

where E_{slow} comprises the twisting, bending and translational energy introduced earlier. The fast modes can be enslaved by any of the slow modes. Since we are mainly looking for a source of positive cooperativity we perform the calculation for fast modes enslaved by opening of the groove f, and then outline the calculation for fast modes enslaved by twisting and bending.

Previously we pointed out that the groove breathing can be decomposed into three principal directions and each of them treated separately. Here we start with the assumption that every fast mode is coupled to the amplitude of the groove width at the site of enslavement and show that the same decomposition can be applied.

In addition, we assume that every fast mode interacts with the slow mode at a single point along the groove (Fig. 5.17). Let us denote fast enslaved modes by the index m where m runs from 1 to M and the global modes are denoted by n where $n = 1, .., N$. There are M points along the groove $s_1, \ldots, s_M$ at each of which a single fast mode interacts with the slow mode.

The individual energy terms under these assumptions are

$$E_{slow} = \int_0^L \left(\frac{1}{4}k(f'^2 + g'^2) + \frac{1}{2}\lambda f^2 \right) \mathrm{d}^3 s, \tag{5.21}$$

$$E_{fast} = \sum_{m=1}^M \left(-V_{f_0} \left(-\frac{k_v \left[f(s_m) \right]^2}{2k_B T} + 1 \right) + \frac{1}{2} \frac{k_f z_m^2}{\exp\left(\frac{k_k [f(s_m)]^2}{2k_B T} \right)} \right), \tag{5.22}$$

where z_m is the m-th fast mode coordinate, k_v and k_k are the coupling constants, k_f is the force constant of the fast modes and V_{f_0} the depth of the potential of unperturbed fast modes. The coupling constants are given without loss of generality the dimension of a spring constant. We assume that the coupling and force constants are identical for each fast mode. These assumptions are taken in order to simplify the calculation and thus demonstrate the effect of the fast modes on the allosteric signaling more clearly and can easily be extended to a more general case.

The partition function of the system is

$$
Z = \iint\limits_{\mathbb{R}^{M+6}} \mathrm{d}f\,\mathrm{d}g\,\mathrm{d}\mathbf{z}\,\exp\left\{-\frac{1}{k_B T}\left[\int\limits_0^L \left(\frac{1}{4}k(f'^2 + g'^2) + \frac{1}{2}\lambda f^2\right)\mathrm{d}s\right.\right.
$$
$$
\left.\left. + \sum_{m=1}^M \left(-V_{f_0}\left(-\frac{k_v\,[f(s_m)]^2}{2k_B T} + 1\right) + \frac{1}{2}\frac{k_f z_m^2}{\exp\left(\frac{k_k[f(s_m)]^2}{2k_B T}\right)}\right)\right]\right\}. \tag{5.23}
$$

Let us first integrate over the fast mode coordinates z_m

$$
Z = \exp\left(\frac{MV_{f_0}}{k_B T}\right)\left(\frac{2\pi k_B T}{k_f}\right)^{(M/2)}\int\limits_{\mathbb{R}^6}\mathrm{d}f\,\mathrm{d}g\,\exp\left\{-\frac{1}{k_B T}\left[\int\limits_0^L\left(\frac{1}{4}k(f'^2 + g'^2) + \frac{1}{2}\lambda f^2\right)\mathrm{d}s\right.\right.
$$
$$
\left.\left. + \sum_{m=1}^M\left(\frac{V_{f_0}k_v}{2k_B T}\,[f(s_m)]^2 - \frac{k_k}{4}\,[f(s_m)]^2\right)\right]\right\}. \tag{5.24}
$$

Assuming that the points s_m are regularly distributed, the sum can be replaced by an integral

$$
Z \propto \int \mathrm{d}f\,\mathrm{d}g\,\exp\left\{-\frac{1}{k_B T}\left[\int\limits_0^L \mathrm{d}s\left(\frac{1}{4}k(f'^2 + g'^2) + \left(\frac{\lambda}{2} + \frac{M}{2}\left(\frac{V_{f_0}k_v}{k_B T} - \frac{k_k}{2}\right)\right)f^2\right)\right]\right\}, \tag{5.25}
$$

where all the constants that cancel in $\Delta\Delta G$ were left out. In order to remove the length dependence we introduce a substitution $s/L = x$. After rearrangement we arrive at

$$
Z \propto \int \mathrm{d}f\,\mathrm{d}g\,\exp\left\{-\frac{\lambda L}{2k_B T}\left[\int\limits_0^1 \mathrm{d}x\left(\frac{\xi^2}{L^2}(f'^2 + g'^2) + \left(1 + \frac{M}{\lambda}\left(\frac{V_{f_0}k_v}{k_B T} - \frac{k_k}{2}\right)\right)f^2\right)\right]\right\}, \tag{5.26}
$$

where we preserved the notation from the previous section $\xi = \sqrt{k/2\lambda}$. In order to ensure integrability of the partition function the factor in front of f^2 has to be positive, posing

a condition on the coupling constants k_k and k_v

$$\frac{M}{\lambda} \underbrace{\left(\frac{V_{f_0} k_v}{k_B T} - \frac{k_k}{2} \right)}_{\eta} > -1, \tag{5.27}$$

where the scaling factor η has been introduced. This condition is fulfilled for

$$k_k < \frac{2 V_{f_0} k_v}{k_B T} + \frac{2\lambda}{M} \tag{5.28}$$

We observe that enslaving of M fast modes rescales the rigidity constant λ by a factor $M\eta$.

The partition function is formally identical to the non-enslaved case and therefore f and g can be decomposed into the three principal components and each of them treated separately. The total allosteric free energy is then the sum of the individual contribution from each component.

We assume that the scaling factor $M\eta$ is independent of the binding state and analyze its effect on the allosteric free energy. Since the enslaving of the fast modes in this way influences only λ and we are interested in positive cooperativity, let us set $\alpha = 1$ for the rest of this section. Depending on the sign of η the rescaling either amplifies or reduces the positive cooperativity.

Case 1. if $\eta < 0$ then the positive cooperativity is amplified. This corresponds to the regime where $k_k > 2 k_v V_{f_0} / k_B T$, i.e. the coupling of the fast potential width to the groove breathing amplitude is stronger than the coupling of the fast potential depth V_{f_0}. Moreover we observe that $\Delta\Delta G$ tends to $-\infty$ for $M\eta \to -\lambda$.

Case 2. if $\eta > 0$ then the positive cooperativity is reduced, and $\Delta\Delta G \to 0$ for $M\eta \to \infty$. That is, if the coupling of the potential depth V_{f_0} to f^2 dominates, the value of the slow allosteric free energy is reduced and for large $M\eta$ tends to 0.

The dependence of the allosteric free energy on the factor $M\eta$ is plotted in Fig. 5.18 for the geometry defined by the example system DNA/H33258 studied in this thesis. The strength of the coupling (parameters k_v and k_k) defines the factor η and thus the character of the enslaving effect. The number of modes amplifies the effect determined by the sign of η. For $M\eta \to -\lambda$ the allosteric free energy can be amplified substantially.
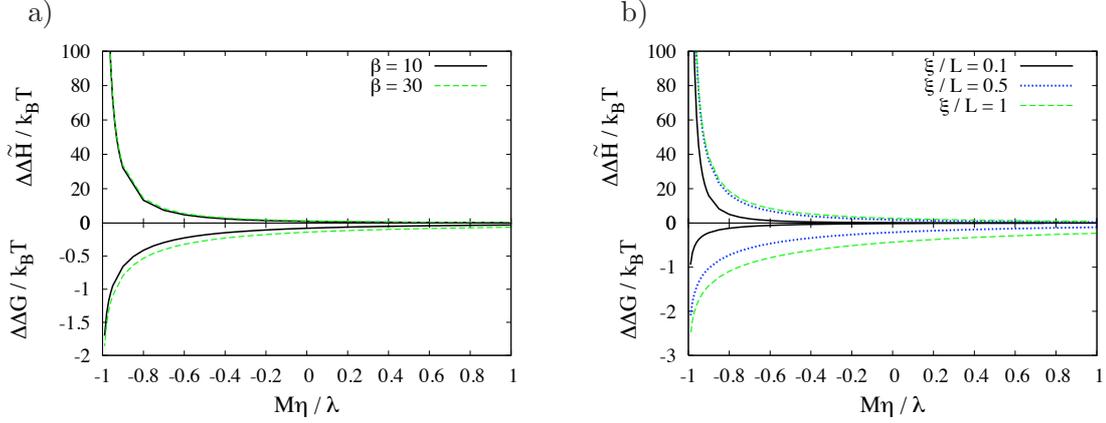
FIGURE 5.18: Enthalpy (top) shown together with the allosteric free energy (bottom) as a function of the scaling factor $M\eta/\lambda$ introduced in Eq. (5.27) displayed for different values of a) β and b) ξ/L. Geometry of the system is defined by the DNA interaction with H33258 ($a = 0.435$, $d = 0.13$). In (a) $\xi = 0.3$, in (b) $\beta = 20$.

The fast modes may be enslaved by any of the slow modes. Let us only outline the calculation in the case of fast modes enslaved by twisting and bending. In the case of twisting, the fast mode does not feel the absolute value of the twist angle but feels the gradient of the angle. This needs to be reflected in the energy expression that reads

$$E_{tw} = \int_0^L \frac{1}{2}C\Theta'^2 \mathrm{d}s + \sum_{m=1}^M \left(-V_{f_0}\left(-\frac{k_v\left[\Theta'(s_m)\right]^2}{2k_BT} + 1 \right) + \frac{1}{2}\frac{k_f z_m^2}{\exp\left(\frac{k_k[\Theta'(s_m)]^2}{2k_BT}\right)} \right). \quad (5.29)$$

Similarly, if bending enslaves fast modes, the important quantity is the local curvature. The energy expression in then

$$E_b = \int_0^L \frac{1}{2}A r''^2 \mathrm{d}s + \sum_{m=1}^M \left(-V_{f_0}\left(-\frac{k_v\left[r''(s_m)\right]^2}{2k_BT} + 1 \right) + \frac{1}{2}\frac{k_f z_m^2}{\exp\left(\frac{k_k[r''(s_m)]^2}{2k_BT}\right)} \right). \quad (5.30)$$

The expressions are formally very similar to the enslaved translation modes. We find that the enslaved modes rescale the twisting and bending rigidity by the factor

$$\eta = \frac{V_{f_0}k_v}{k_BT} - \frac{k_k}{2}. \quad (5.31)$$

The parameters in principle depend on the type of the slow mode. In chapter 4 we

have demonstrated that bending and twisting can only result in negative cooperativity. Enslaving of the fast modes can lead to amplification or reduction of this allosteric free energy depending on the scaling factor:

Case 1. If $\eta < 0$ the negative cooperativity is amplified; $\Delta\Delta G$ tends to ∞ for $M\eta \to -A$ or $M\eta \to -C$, respectively. This corresponds to the regime where $k_k > 2k_v V_{f_0}/k_B T$, i.e. the coupling of the fast potential width to the twist gradient or bending curvature is stronger than their coupling to the fast potential depth V_{f_0}.

Case 2. If $\eta > 0$ the anti-cooperativity is reduced, and $\Delta\Delta G \to 0$ for $M\eta \to \infty$. That is, if the coupling of the potential depth V_{f_0} to θ'^2 or r''^2 dominates, the value of the slow allosteric free energy is reduced and for large $M\eta$ tends to 0.

Although positive cooperativity cannot be generated by enslaving of the fast modes, the negative cooperativity can be suppressed.

If fast modes are enslaved by all three types of slow modes the slow part of the energy expression is a sum of the previously derived terms (4.1) and (5.9). The fast part of the energy expression is

$$E_{fast} = \sum_{m=1}^{M} \left[-V_{f_0} \left(-\frac{k_{v,tw}\left[\Theta'(s_m)\right]^2}{2k_B T} - \frac{k_{v,b}\left[r''(s_m)\right]^2}{2k_B T} - \frac{k_{v,trans}\left[f(s_m)\right]^2}{2k_B T} + 1 \right) \right.$$
$$\left. + \frac{1}{2}\left(e^{-\frac{k_{k,tw}\left[\Theta'(s_m)\right]^2}{2k_B T}} + e^{-\frac{k_{k,b}\left[r''(s_m)\right]^2}{2k_B T}} + e^{-\frac{k_{k,trans}\left[f(s_m)\right]^2}{2k_B T}} \right) k_f z_m^2 \right] \qquad (5.32)$$

The standard statistical mechanics calculation results in independent free energies for the twisting, bending and groove breathing with their effective spring constants rescaled by the respective factor $M\eta$.

As has been shown before the free energy involving the enslaved fast modes is composed of compensating enthalpy and entropy contributions (Sec. 2.5). We calculate enthalpy as

$$H = -\frac{\partial \ln Z}{\partial \beta}, \qquad (5.33)$$

where $\beta = 1/k_B T$. The partition function is obtained by evaluating the integral in Eq. (5.26) and results in

$$Z = \left(\frac{\pi}{\beta}\right)^{\frac{3M}{2}} \left(\frac{2}{k_f}\right)^{\frac{M}{2}} \frac{\exp\left(M V_{f_0}\beta\right)}{\left(|\mathbf{M}_f| |\mathbf{M}_g|\right)^{1/2}} \tag{5.34}$$

where the matrix $\mathbf{M}_f$ is associated with the normal modes of f, and $\mathbf{M}_g$ with g. For the sake of clarity let us denote all matrices in this subsection in with boldface capital letters. We are only interested in the allosteric enthalpy $\Delta\Delta H$ and can therefore neglect all terms that remain constant during the binding events. Of the remaining terms only $\mathbf{M}_f$ is a function of β and the enthalpy is thus

$$H = \text{const.} \frac{1}{2|\mathbf{M}_f|} \frac{\partial |\mathbf{M}_f|}{\partial \beta}. \tag{5.35}$$

Derivative of a determinant can be calculated using the Jacobi formula [218]

$$\frac{d|\mathbf{M}|}{dx} = \text{tr}\left(\text{adj}(\mathbf{M}) \frac{d\mathbf{M}}{dx}\right), \tag{5.36}$$

where $\text{adj}(\mathbf{M})$ denotes the adjugate of matrix $\mathbf{M}$. If $\mathbf{M}$ is invertible then

$$\text{adj}(\mathbf{M}) = \mathbf{M}^{-1}|\mathbf{M}| \tag{5.37}$$

and

$$\frac{d|\mathbf{M}|}{dx} = |\mathbf{M}| \text{tr}\left(\mathbf{M}^{-1} \frac{d\mathbf{M}}{dx}\right). \tag{5.38}$$

And finally

$$H = \frac{1}{2} \text{tr}\left(\mathbf{M}_f^{-1} \frac{\partial \mathbf{M}_f}{\partial \beta},\right). \tag{5.39}$$

Because only the diagonal terms of $\mathbf{M}_f$ are dependent on β, we can simplify the expression for H to

$$H = \frac{1}{2} \text{tr}\left(\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^{-1} M \frac{V_{f_0} k_v}{\lambda} \mathbf{I}\right) \tag{5.40}$$

where $\mathbf{P}$ is an invertible matrix such that $\mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^{-1} = \mathbf{M}_f$ and $\mathbf{\Lambda}$ is a diagonal matrix. The factor $M V_{f_0} k_v/\lambda$ is found from differentiating $M\eta/\lambda$ with respect to β and $\mathbf{I}$ is an identity matrix. After some rearranging we find that

$$H = M \frac{V_{f_0} k_v}{2\lambda} \sum_{m=1}^{M} \frac{1}{\lambda_m}, \tag{5.41}$$

where λ_m are the eigenvalues of the matrix $\mathbf{M}_f$. This approach has the advantage that the corrupted eigenvalues arising from the truncation of the matrix $\mathbf{M}_f$ can be disregarded in the same way as during the $\Delta\Delta G$ calculation.

We observe that $H = 0$ if $k_v = 0$, that is when there is no coupling of fast modes to the depth of the slow potential. When $k_v > 0$ then $\Delta\Delta H > 0$ and increases with $M\eta/\lambda$ approaching -1 from the right (Fig. 5.18). The factor in front of the sum in Eq. (5.41) remains constant during the sequential ligand binding, so we therefore plot

$$\Delta\Delta\tilde{H} = \sum_{m=1}^{M} \frac{1}{\lambda_m^{free}} - 2\sum_{m=1}^{M} \frac{1}{\lambda_m^{1:1}} + \sum_{m=1}^{M} \frac{1}{\lambda_m^{2:1}} \tag{5.42}$$

as a function of the factor $M\eta$. The magnitude of $\Delta\Delta\tilde{H}$ is scaled by $MV_{f_0}k_v/(2\lambda)$ while the sign remains the same assuming $k_v > 0$. In Fig. 5.18, $\Delta\Delta\tilde{H}$ is plotted alongside $\Delta\Delta G$ to demonstrate that $\Delta\Delta G$ is composed of compensating entropic and enthalpic part and the positively cooperative effect is entropy driven.

Interestingly $\Delta\Delta H > 0$ even when $\eta = 0$. In this case the allosteric free energy equals the allosteric free energy of the slow modes only but is composed of compensating enthalpic and entropic parts. This is a generalized version of the case $A = 0$ discussed in relation with the homodimers in chapter 3.

5.3.1 Fast Modes in the DNA/Hoechst 33258 System

Experiments and MD simulations of the DNA/H33258 system have evaluated the allosteric free energy $\Delta\Delta G \lesssim -7k_B T$. We identified groove breathing as a source of positive cooperativity, but its contribution to $\Delta\Delta G$ appears to be small ($\sim -0.5\,k_B T$). Furthermore, in the previous chapter, we estimated the contribution from twisting and bending to be of the same magnitude, but opposite sign, yielding the overall $\Delta\Delta G \approx 0\,k_B T$. In this section we have investigated the potential effect of enslaving fast, localised modes and found that they can amplify positive allostery if enslaved to the groove breathing modes.

We suggested base motions as candidates for localised, enslaved modes and measured the effect of H33258 on the amplitude of their fluctuations (Fig. 5.16). Let us now estimate the potential contribution of these modes to allosteric free energy. A fast, localised mode can amplify positive cooperativity only if it is enslaved by the groove breathing and the

base pair parameter	affected base pairs	$\frac{k_{apo}}{k_{2:1}}$
Shear	3-5 & 8-10	0.86
Stagger	3-4 & 9-10	0.81
Buckle	2-4 & 9-11	0.64
Propeller	3 & 10	0.73

TABLE 5.1: The list of base pair parameters that are affected by H33258 binding. The ratio $k_{apo}/k_{2:1} = \langle z_{2:1}^2 \rangle / \langle z_{apo}^2 \rangle$ was obtained by measuring the RMSD of the particular parameter in the apo and 2:1 complex in the MD simulations.

coupling factor $\eta < 0$. The coupling factor is negative only if the amplitudes of the slow and fast mode are correlated, i.e. a reduction of the amplitude of the slow mode is reflected in reduction of amplitude of the fast mode. For this reason we focus on shear, stagger, buckle and propeller and assume they are coupled to the groove width f.

The presence of each molecule of Hoechst 33258 affects shear on 3 bases, stagger on 2 bases, buckle on 3 bases and propeller on 1 base. These form a set of 18 localised modes, which we consider enslaved by the groove breathing. We now estimate an average factor η/λ (Eq. (5.27)) for these modes. Firstly, we do not have any means to measure the coupling of the potential depth of the fast modes V_{f_0} to the groove width and so assume it is negligible and set $k_v = 0$. The scaling factor now simplifies to $\eta = -k_k/2$.

We further assume that k_k is not affected by the H33258 binding and that the potential of each fast enslaved mode has a form

$$V_{f_i} = \frac{1}{2} \frac{k_f z_i^2}{\exp\left(\frac{k_k f_i^2}{2k_B T}\right)} \equiv \frac{1}{2} k_i z_i^2, \tag{5.43}$$

where the index i refers to the binding state of the DNA, and k_i is the observed force constant of the fast mode. We then extract the ratio $k_{apo}/k_{2:1}$ from the RMSDs of the fast modes presented in Fig. 5.16. These ratios are very similar for all modes (Table 5.1) and we can thus use the simplifying assumption of identical coupling constant for all fast modes. The value of $k_{apo}/k_{2:1}$ is taken as the average value of all the ratios, which yields $k_{apo}/k_{2:1} = 0.8$.

The groove width f has been measured during the course of the simulation; the RMSD values along the helical axis in three binding states are presented in Fig. 5.6. Averaging over the helical axis in each state yields $\langle f_{apo}^2 \rangle = 2.9\,\text{Å}^2$ and $\langle f_{2:1}^2 \rangle = 1\,\text{Å}^2$.

A value of $k_k = 0.24\,k_BT\,\text{Å}^{-2}$ is then obtained by comparing the potential (5.43) in the apo and 2:1 state and inserting the measured values of f_i^2 and k_i.

In order to estimate λ we again use the groove width fluctuations (Fig. 5.6). The equipartition theorem then gives

$$\lambda = \frac{3k_BT}{2\Delta s\,\langle f^2\rangle}L, \tag{5.44}$$

where $\Delta s = 3.4\,\text{Å}$ is the distance between base pairs, $L = 12\Delta s$ and $\langle f^2\rangle = 2.9\,\text{Å}^2$ is the mean square deviation of the groove width obtained as an average over all measuring locations in Fig. 5.6. The factor of three accounts for three degrees of freedom. This method is somewhat crude because it neglects the extensibility of the groove edges. We find that $\lambda = 6\,k_BT\,\text{Å}^{-2}$. Inserting $M = 18$ (9 in each binding site) enslaved modes results in the value of the scaling parameter

$$\frac{M\eta}{\lambda} = -\frac{Mk_k}{2\lambda} = -0.36. \tag{5.45}$$

Inserting this value into the expression for allosteric free energy with previously estimated parameters $a/L = 0.435$, $d/L = 0.13$, $\xi = 3-4\,\text{bp}$ and $\beta = 10-30$ results in $\Delta\Delta G$ ranging between -0.1 k_BT and -0.23 k_BT per degree of freedom, hence between -0.3 k_BT and -0.7 k_BT for the whole system. This corresponds to approximately 40 % increase compared to the nonenslaved allosteric free energy.

These values are still relatively low for a biological system. However we observe in Fig. 5.18 that larger values of $M\eta/\lambda$ can give rise to significantly larger free energies. The value of $M\eta/\lambda = -0.36$ may be more negative for several reasons. Firstly, we only examined the base pair fluctuations but there may be many more fast modes enslaved which would result in larger M. Secondly, the estimate of λ is only crude and due to neglecting of the groove edge stiffness k is probably overestimated.

If $M\eta/\lambda = -0.9$, corresponding to physically reasonable set of parameters such as $k_k = 0.24\,k_BT\,\text{Å}^{-2}$, $\lambda = 5\,k_BT\,\text{Å}^{-2}$ and $M = 36$, then for $\beta = 30$, $\xi = 4\,\text{bp}$ and three degrees of freedom $\Delta\Delta G = -2.5\,k_BT$. Another possible source of larger allosteric free energy is the decay length ξ. Table 5.2 summarizes values of $\Delta\Delta G$ arising from groove breathing with fast modes enslaved for several combinations of $M\eta/\lambda$, decay length ξ and β and again shows that values around $3\,k_BT$ can be achieved for physical values of parameters.

	$M\eta/\lambda$	$\Delta\Delta G_{3D}/k_BT$	
		$\beta = 10$	$\beta = 30$
$\xi = 3\,\mathrm{bp}$	-0.4	-0.3	-0.5
	-0.9	-1.6	-1.9
$\xi = 4\,\mathrm{bp}$	-0.4	-0.5	-0.7
	-0.9	-2.0	-2.5
$\xi = 5\,\mathrm{bp}$	-0.4	-0.7	-1.0
	-0.9	-2.4	-2.9

TABLE 5.2: Allosteric free energy of the groove breathing with fast modes enslaved. The values are given for three degrees of freedom and the fast mode coupling is defined by the factor $M\eta/\lambda$. Two values of $M\eta/\lambda$ are presented: 1) $M\eta/\lambda = -0.4$ has been found from the simulations, 2) $M\eta/\lambda = -0.9$ illustrates how biologically relevant values of allosteric free energy can be generated by the model. Example values of parameters yielding $M\eta/\lambda = -0.9$ are $\{k_k = 0.24\,k_BT\,\text{Å}^{-2},\ \lambda = 5\,k_BT\,\text{Å}^{-2},\ M = 36\}$.

The experimentally measured value of the allosteric free energy was $-6.8\,k_BT$ and simulations yielded $-12\,k_BT$. We have shown that the mechanisms presented in this thesis can be responsible for about a third of the experimental value.

We have shown that the enthalpy arising from enslaved fast modes is equal to 0 if $k_v = 0$, but for $k_v > 0$ the allosteric enthalpy $\Delta\Delta H > 0$. Harris *et al.* have measured $\Delta\Delta H \simeq 2\,k_BT$, and therefore acting anti-cooperatively. We have demonstrated that such a positive allosteric enthalpy can arise from enslaved fast modes. There are three possible sources of the positive enthalpy: 1) the coupling factor k_v can be positive for the modes enslaved by the groove breathing and thus slightly reducing the amplification of the positive cooperativity whilst giving rise to positive $\Delta\Delta H$, 2) k_v can be positive for fast modes enslaved by bending and/or twisting and thus reducing the anti-cooperativity and at the same time giving rise to the positive allosteric enthalpy or 3) the enthalpy can result from unfavourable interactions between DNA and the ligand.

5.4 Conclusions

The aim of this chapter was to find a potential origin of the positive cooperativity observed upon sequential binding of ligands to DNA, in particular when the ligands are small minor groove interacting molecules. A number of small minor groove binding ligands display positive cooperativity without conformational change when binding

sequentially to short DNA oligomers [101, 157, 159, 170, 171]. The similar thermodynamic properties, measured independently of the structural details of the oligomer or the ligand, suggest that one mechanism is responsible for the allosteric effects. One of these systems has been studied in more detail than the others, namely Hoechst 33258 binding to DNA oligomers [101, 155, 171]. We constructed a general model of allosteric DNA oligomer and parameterised and tested it from the available simulations of the Hoechst/DNA system.

In the previous chapter we have shown that binding of Hoechst locally modifies the bending and twisting rigidities of the oligomer, but that this can only result in negative cooperativity. In particular, we have estimated the negative cooperativity arising from this phenomenon to be $\Delta\Delta G_{twist+bend} = 0.5\,k_B T$ for the Hoechst/DNA system studied by Harris *et al.* [101].

We then discussed a number of possibilities for sources of positive cooperativity in DNA and decided to focus on DNA groove breathing. It has been shown that the minor groove-binding ligands perturb the natural groove breathing [101], and here we have demonstrated using a model with appropriate coarse-graining that such perturbation can propagate along the DNA and contribute to the allosteric effects.

We evaluated the extent of the contribution and found that it depends on 5 parameters: 3 elastic (the decay length ξ, the phonon stiffening α and the groove stiffening β) and 2 geometrical (the ligand length a and the ligand spacing d). The normalised decay length ξ/L denotes the distance over which the allosteric signal decays by the factor e and L the length of the rod. The groove breathing in our model is represented by two nearly independent processes - stretching of the groove edges controlled by the multiplicative factor α and opening of the groove controlled by β. Positive cooperativity was found to arise only for values of α near unity and to be maximal for $\alpha = 1$.

The dependence on the geometry of the system is more complex than in the case of twisting and bending thanks to the additional factor: the decay length ξ. In the majority of cases the system's cooperativity increases with the length of the ligand, with decreasing spacing between the ligand and with increased decay length. However for small values of a and d compared to the length of the rod, this is reversed and systems with smaller decay length become more cooperative. This behavior is illustrated in Fig. 5.11 and 5.12.

The parameterisation of the groove breathing for the DNA/Hoechst system revealed that some positive cooperativity arises, but is approximately of the same size as the

negative cooperativity coming from twisting and bending, resulting in a cancelation of the two effects.

The last potential source of positive cooperativity investigated was the enslaving of fast modes. As explained before, fast modes, despite being localized, can couple to the slow, global modes and assist cooperativity. It has been shown that the enslaving can either amplify or reduce cooperativity but cannot change its type from positive to negative or vice versa [42]. Substantial increase in the positive cooperativity can therefore only occur if the fast modes are enslaved by the groove breathing. We performed the calculation and found that, in agreement with [42], amplification occurs in the regime where coupling of the fast mode potential width to the groove breathing amplitude is much stronger than the coupling of the fast mode potential depth to the amplitude. Positive cooperativity can be amplified considerably for a large number of modes and/or strong coupling. Moreover, if fast modes are enslaved by twisting and bending in the alternative regime, in which the slow mode mainly influences the depth of the fast potential, the negative cooperativity can be effectively reduced.

In an analysis of the DNA/Hoechst simulations, 18 fast modes emerged as potentially enslaved by the groove breathing. We evaluated the consequent allosteric free energy and found value by $\sim 40\,\%$ to $-0.7\,k_BT$. This constitutes a noticeable increase but the value is still low compared to the experimental and computational results. We then speculated that under- or overestimation of individual constants and more harnessed fast modes could lead to biologically relevant values of allosteric free energy of $\approx -3\,k_BT$. This, after the subtraction of $0.5\,k_BT$, corresponds to a third of the experimentally measured value of $\Delta\Delta G = -6.8\,k_BT$.

Although significant allosteric free energy can be obtained through enslaving of fast modes, it is plausible that other mechanisms take part in the allostery of the Hoechst/DNA system and possibly any allosteric system involving DNA. We have considered all potentially significant harmonic contributions. This leaves anharmonic motions as a next most probable source. Anharmonic effects arise when large amplitudes of the DNA motion are considered and result in nonlinear equations describing the motion.

Examples of such models are given in [154] and generally lead to soliton solutions, a type of travelling wave or more precisely a pulse. Any perturbation of the DNA structure can thus initiate a formation of a soliton that propagates along the DNA [154, 219]. It may be possible that the wave somehow propagates the allosteric signal and increases

affinity for the second molecule. The soliton propagation itself is however a matter of debate and remains to be found in experiments and/or simulations. Other mechanisms involving anharmonic motions may play a part in allosteric effects though.

Chapter 6

Conclusions and Future Work

The importance of conformational dynamics for the function of biological molecules has become increasingly recognised. In this thesis we investigated its role in allosteric regulation with a set of coarse-grained models. The aim was to draw on the initial work of Hawkins and McLeish and improve the understanding of the phenomenon of dynamic allostery through developing the existing and constructing new models. The construction of additional models allowed us to gain a better perspective and discover common aspects and principles of this ubiquitous phenomenon.

The coarse-grained models presented in this thesis are very crude and were built with no ambition to represent the full complexity of real biomolecules. Detailed atomistic simulations can be used to reproduce complex phenomena such as protein-ligand interactions, but I believe that simple models are essential for true understanding. We therefore focused on keeping the models intuitive and as simple as possible yet able to capture the allosteric effect of interest and reveal its fundamental principles. Simultaneously, we attempted to design models resembling the real systems closely enough to allow parameterisation from experiments and simulations.

The strategy used in the coarse-graining process was to capture the most intuitive dynamic modes of the molecule and investigate their ability to give rise to allostery. In both examples discussed the initial model was not sufficient to describe the full extent of the observed allosteric effect and consequently was extended to include additional, selected, features of the real system. This allowed the elucidation of the role of different structural vibrations and the role of fluctuations on different timescales.

The results were divided into three chapters, one concerned with a homotropic allostery in proteins with two and more identical subunits, and two addressing allostery propagation through elastic rod like molecule, such as DNA.

In Chap. 3 we focused on homotropic allostery in homodimers, i.e. cooperative binding of two identical ligands to proteins composed of two identical subunits. We constructed a coarse-grained model of this system by gradually increasing its level of complexity; beginning with a model that included only one breathing mode per subunit and elastic coupling between them. Even at this level, the model demonstrated the possibility of propagation of allosteric signal purely via slow global motions. We investigated the role of elastic coupling between the monomers and concluded that its changes upon ligand binding are crucial to the type of the allosteric effect. In order to recover experimentally observed values of thermodynamic parameters and to comprise some of the complexity of the real proteins we extended the model to include additional number of global modes and also localised fast modes. These modes are naturally present in proteins and it is expected that they take part in the allosteric signaling too. Increasing the number of global modes resulted in an amplified allosteric signal, leading to biologically relevant allosteric free energies. Fast modes were included in the model as coupled to the slow global modes and were shown to not to alter the allosteric free energy substantially but to split the previously purely entropic effect into finite entropic and enthalpic contributions.

We validated this model on a test case homodimer: the CAP protein. CAP is the first protein displaying allostery without conformational change to be studied experimentally in detail and as such has a great exemplary power in the biological community. Our model demonstrated that changes in global and local vibrational modes of the protein can give rise to the allosteric communication observed in experiments. Furthermore we showed that the protein is using the dynamics in an optimized manner, which serves as yet another evidence that the widely accepted structure-function paradigm needs to be extended to include dynamics.

We then proceeded to show that the conclusions obtained from the model of a homodimer may represent general phenomena by extending the model to include a larger number of protein units. This extended model demonstrated that the same mechanism of allostery can be exploited in higher oligomers too, but also indicated that other phenomena may come into play in larger systems. These have not been studied in this thesis and are mentioned as a part of the future work in the next section.

Chapters 4 and 5 addressed allostery in a very different biomolecule, namely DNA. We chose DNA as a model system because mechanical properties of DNA are relatively well known. Concretely, the properties of DNA at the length scales relevant for allostery are best represented by the elastic rod model. This model differs from the previously studied models of proteins by a large number of spatially delocalised modes. Additionally, we used a different method to model ligand binding. The attachment of a substrate was assumed to modify the rigidity of the rod along a stretch of finite length. In contrast to the previous discrete models studied by ourselves and Hawkins and McLeish [38, 41, 42], this type of interaction gives rise to normal modes with altered structure.

The altered structure of normal modes is a consequence of coupling of modes near in the frequency spectrum and is responsible for the negative cooperativity observed upon the sequential identical ligand binding. The local reduction in amplitude of spatial fluctuations gives rise to increased amplitude in the surrounding regions. This then implies larger entropic cost of binding to the vicinity of the first ligand. We have already observed a similar effect in the model of homodimer, in which we introduced the simplest level of elastic coupling. In this case, the first ligand binding in the negatively cooperative binding sequence was accompanied by increased amplitude of fluctuations in the unliganded subunit. This effect has however not been observed experimentally and it remains to be verified if it constitutes a common mechanism of negatively allosteric systems.

Choosing DNA as a model system has a drawback. Allosteric effects in DNA, despite clearly present, have not been subjected to much experimental investigation. Binding of proteins up- or downstream from the transcription site is known to trigger or inhibit transcription, which strongly indicates that allosteric effects, positive and negative, are at play. Because of their size (tens to hundreds of base pairs in addition to the proteins), these systems are however almost impossible to investigate directly with current experimental techniques. The mechanism of communication between the interaction sites is therefore disputed and is generally ascribed to DNA looping or other structural modifications [210].

Attempts to study allosteric signal propagation through DNA directly have been made on smaller systems and strongly indicated the importance of dynamics [101, 171]. These smaller systems however all display positive cooperativity but also share a common binding motif: a tight fit of a long narrow ligand into the minor groove of DNA. This suggests that the minor groove binding drugs primarily explore a different mechanism

for the allosteric communication, associated with more detailed structural fluctuations of DNA. We however believe that the negative dynamic allostery predicted by our calculations is present in nature and that future investigations will identify examples of such anti-cooperative systems.

In order to address the question of positively cooperative minor groove binding drugs we extended the elastic rod model to include minor groove breathing (Chap. 5). We modeled the breathing as a fluctuation of two parallel elastic helices coupled by a harmonic potential. The ligand was assumed to bind between the helices and locally modify their rigidity and the potential between them. We found that two sequential binding events can display positive cooperativity only if the stiffness of the helices, or groove edges remains nearly or completely unmodified and the groove potential stiffens or softens. When evaluated for the test case of the drug Hoechst binding to a short DNA oligomer we found that the allosteric free energy arising from this effect approximately cancels the anti-cooperative effect originating from modified bending and twisting rigidities. One more possible source of positive cooperativity investigated in this thesis was enslaving of fast localised modes. We found that fast modes can amplify the positive cooperativity.

It is possible that other mechanisms, such as anharmonic vibrations, easier displacement of groove-bound water molecules from the second binding site or electrostatic interactions give rise to additional contributions to the allosteric free energy. Association of a ligand with a minor groove is known to be usually entropy-driven, due to the removal of well ordered water molecules from the binding site [220]. It is thus possible that first ligand binding facilitates the hydration of the second binding site, resulting in entropically more favourable association. However, Harris *et al.* studied this possibility for the case of Hoechst and concluded that it is improbable [101]. Electrostatic interactions also need to be considered. In the example taken here the two Hoechst ligands face each other with positively charged piperidine rings separated by a 15 Å wide, mainly water filled, gap. The water screens the repulsive charges, but communication could occur through the low dielectric medium of DNA.

One important conclusion that follows from this work is that positive homotropic cooperativity can only arise through dynamic mechanisms if the amplitudes of fluctuations at two binding sites are correlated, i.e. if a reduction of the amplitude at one binding site results in reduced amplitude on the other binding site. If, on the other hand, the amplitudes at the binding sites are anti-correlated, as we have seen in the case of DNA, then the allosteric effect is negative. This second effect is more likely to occur in more

uniformly elastic molecules, where a larger number of modes can become coupled. It will however occur to certain extent in all biological molecules, which leads us to a conclusion that dynamics naturally supports negative cooperativity. Positive cooperativity can arise, but the molecular design has to be rather specific to allow the change of *frequency* to be dominant over the change of *structure* of the global modes.

This finding underlines the ingenuity with which nature designed biological macromolecules, and in particular proteins, to include discrete regions of high and low elastic modulus. This ensures a certain degree of robustness of the low frequency motions and thus enables proteins to harness these for allosteric signaling of any type.

One of the main challenges for the future will thus be determining what makes certain proteins so special that they are able to propagate the signal for positive cooperativity through dynamic mechanisms.

6.1 Future Work

The combination of analytical statistical mechanics models with computational and experimental parameterisation has proved a powerful technique in extending our understanding of the role of dynamical processes in allosteric signaling and more broadly in biomolecular function. In future we would like to use the findings and observations acquired through the current coarse-grained models and their extensions for a development of a more systematic method of coarse-graining and for identifying general design rules of allosteric systems. The identification of such rules would have great implications for drug development and protein engineering.

The subject of dynamic allostery is relatively unexplored and thus offers many options to improve and expand the work presented here. In this section I focus only on the immediate improvements and natural extensions of our models, which I hope to implement in the near future.

6.1.1 Validation of the Homodimer Model

The homodimer model developed in this work helped to explain dynamic allostery in CAP and also to make several predictions. In order to verify these hypotheses experimental tests are required. These can generally be of two types. Firstly, an extended set

of experiments can be performed on CAP itself, and secondly, the model can be applied to other homodimers. Let me suggest several potential means of testing our predictions. The results of these tests can then be used in a feedback loop to improve or modify our model.

The main prediction of our model is the shape of the free energy landscape (Fig. 3.4). I believe that, in a close collaboration with a group of experimental biophysicists, we could identify and perform a series of point mutations in CAP that would simulate a motion on the landscape. In particular, we predicted that the coupling between the CAP subunits becomes weaker upon each cAMP binding ($\alpha < 1$), which implies that any change in parameter β, i.e. change in internal monomer stiffness upon binding should have only negligible effect on the allosteric free energy. Any change in the value of parameter K_c on the other hand should be felt relatively strongly.

A more ambitious mean to test our predictions would be an attempt to switch the allosteric behaviour of CAP to become positively cooperative by the means of modifying of the elastic parameters. The most straight-forward route would be to change the value of α (the coupling enhancement factor) from smaller than unity to larger than unity. We hope to identify suitable mutations and to test them in collaboration with biophysicists and biochemists in the near future.

Many homodimeric proteins display allosteric effects, but ideally we would like to test the model on other examples of entropy driven allostery. One such system was mentioned in the introduction to Chap. 3, the GCT enzyme. Structural measurements have not identified any conformational change accompanying the negative cooperativity of CTP ligand binding to GCT (fig. 3.2). The NMR relaxation experiments suggested that significant rigidification occurs upon the second CTP binding but were not able to sufficiently characterise the change in dynamics accompanying the first ligand binding. It would be interesting to see if the loosening tightening effect is a general mechanism of dynamic negative allostery in homodimers.

6.1.2 Improvements to the Higher Oligomer Model

There is much room for improvement and generalisation of the model of higher oligomers studied in Sec. 3.4. Let us start discussing the work required on the model of allosteric homotetramers.
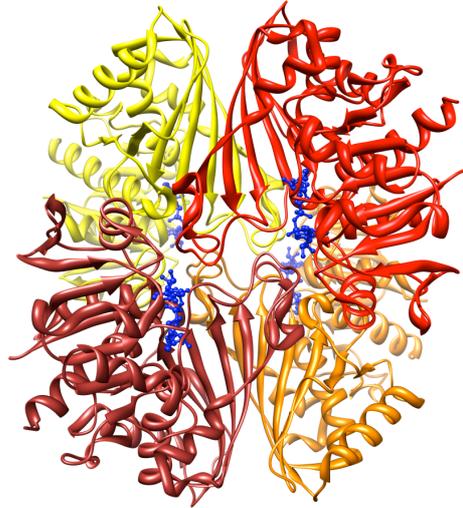
FIGURE 6.1: Ribbon diagram of the fully liganded form of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). The four subunits of GAPDH are colored in different shades of red, the ligand NAD is shown in balls and sticks in blue. PDB code: 1ZNQ.

We have constructed the most basic model of one internal mode assigned to each monomer. The development of the model into the complexity achieved for the model of homodimer is a natural next step and would be required for a credible comparison with experimental results. Furthermore, the confrontation of the model with hemoglobin uncovered a possible complication when extending the dimer model to larger oligomers. For any protein composed of more than two identical subunits the interaction between the n-th and $(n-1)$-st and between the n-th and $(n+1)$-st neighbour can be substantially different. For example, in hemoglobin 35 residues take part in the interaction on the interface between the subunits of the same dimer ($\alpha_1\beta_1$ interface) and only 19 residues are in contact with the neighboring subunit belonging to the other dimer ($\alpha_1\beta_2$ interface). Although a possible extension of the model to include two different coupling constants would entail the complications of additional free parameter, it may be required for a realistic model.

Introducing new parameters can be avoided by studying proteins with comparable coupling at each subunit interface. A tetramer that appears to have relatively balanced interactions between all subunits, GAPDH (Fig. 6.1), has been briefly discussed in Sec 3.4.2. This homotetramer is also interesting for other reasons. It displays a compelling allosteric behavior, namely, GAPDH extracted from yeast binds 4 identical

effectors called NAD with positive cooperativity [221] and GAPDH extracted from muscle binds the same molecules anti-cooperatively [143, 222]. The negatively cooperative GAPDH is a well studied protein, one of the very few, for which structures of the apo [222], partially liganded [143] and fully saturated forms [223] are known. Moreover, the partially liganded structure reveals absence of structural change in the unliganded subunits [143] and thus indicates that dynamic allostery may play an important role in the allosteric binding.

Each of these properties individually renders GAPDH interesting enough for comparison with our tetramer model. We would therefore like to use the experimental values extracted from the literature for the parameterisation of the extended homotetramer model. Most importantly, we hope we would be able to shed light on the reasons for different type of allostery associated with the two forms of GAPDH.

As explained in the main section, the number of binding pathways grows exponentially with the number of subunits in the protein. The four distinct pathways of a pentamer can still be investigated with the methods employed for dimers and tetramers. We would like to subject the model of a pentamer to the same degree of study as the tetramer model. A relevant biological example is also readily available in the form of the B_5 subunit of the AB_5 class of cytotoxins.

The AB_5 class of cytotoxins includes for example the dangerous cholera and Shiga toxins. These toxins are composed of two subunits, the enzymatically active A subunit and the B_5 subunit providing oligosacharide recognition. The homopentameric B_5 subunit can be targeted in therapeutical applications and understanding of the oligosacharide recognition is thus of high importance.

Yung *et al.* employed NMR to study binding of the P^k dimer, an analogue of the natural oligosacharide acceptor globotriaosylceramide, to the pentamer [225]. The P^k dimer consists of two binding domains connected by a flexible linker, each of which binds to a subunit of the pentamer (Fig. 6.2 b). They reported that the free pentamer undergoes large intersubunit fluctuations which become suppressed upon the first P^k binding. The stabilisation is reflected in large entropic cost of binding, but leads to reduced entropic penalty for the subsequent bindings. Apart from displaying dynamic allostery, this protein is also characterised by identical interaction between all of its subunits (Fig. 6.2).

Oligomers composed of more that five subunits will have to be tackled with a slightly different approach. Instead of studying and comparing all possible pathways we would
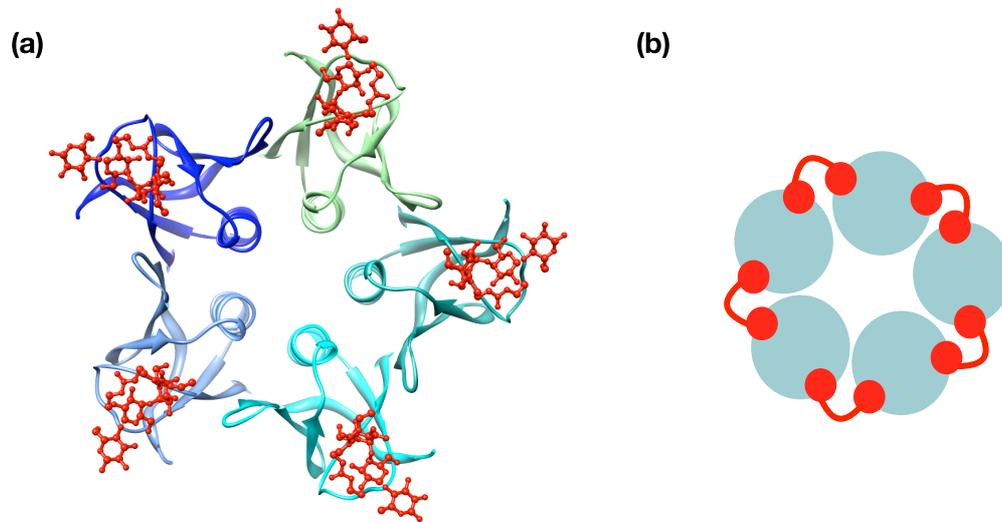
FIGURE 6.2: Two binding modes of the homopentamer B_5. (a) Ribbon diagram of 3d structure of the pentameric B_5 subunit of the AB_5 toxins. Here shown bound to the STARFISH inhibitor (in red). PDB code: 1QNU. (b) A cartoon of the B_5 pentamer bound to five molecules of the P^k dimer. Each dimer is believed to straddle adjacent monomers [224].

like to identify the best strategies of binding, i.e. the entropic cost of nearest neighbor binding versus distant binding. We would also like to test the hypothesis that the increased number of subunits leads to more continuum like behavior, which we investigated in connection with the DNA, and should thus result in preferentially negative cooperativity.

Appendix A

Computational Analysis of CAP Dynamics

In this Appendix we summarize the results obtained from computational analysis of the three dimensional structure of the catabolite activator protein (CAP). The objective of this study was to model slow modes of CAP using coarse-grained computer simulations. We expected to obtain information on number and structure of rigid and flexible domains and their evolution during the sequential cAMP binding. We were in particular interested if the loosening-tightening effect observed in experiments [4] could be recovered. The analysis was performed using a selection of elastic network models implemented at web servers iGNM, ElNemo and WEBnm and with a combination of programs FIRST and FRODA implemented at server Flexweb. The individual techniques were described in detail in chapter 2.

All simulations presented here were performed on the truncated version of CAP (CAP^N, residues 1 - 138) used by the group of Kalodimos for their experiments [4]. The coordinates of the protein were obtained from the crystal structure of the doubly liganded full-length protein (pdb entry 1G6N) by selecting desired residues and stripping off cAMP ligands for the singly liganded and unliganded versions. The simulations were repeated on minimized structures of apo-CAP and singly liganded CAP, but the differences found were negligible.

iGNM

In the Gaussian network the protein is modeled on the single site per residue basis. We modified the structure pdb file so that a site was also located in the middle of each cAMP ligand. The model also included constraints imposed by the nearest neighbors along the backbone. iGNM server calculates 20 slowest, 20 fastest modes, correlation between fluctuations and performs several other analyses (for details see ref. [112]).

Analysis of 20 slowest modes of all three ligation states shows that CAP^N becomes stiffer upon each cAMP binding. A list of first six eigenvalues is shown in Table A.1. A graphical representation of the modes indicates that each monomer of CAP^N is divided into two rigid domains: one is formed by the central helix and the other by the β−sheet complex (Fig. A.1). The analysis was performed for two different cutoff values: 7.3 Å and 10 Å. The value of 7.3 Å is used by the authors of GNM database of motions of all known protein structures, because it yields the best agreement between theoretical and experimental B-factors [112]. The cutoff distance of 10 Å is default at the website. It has been shown that the choice of the cutoff distance has little effect on the protein dynamics [112], however the fluctuation amplitudes are inversely proportional to the number of contacts per residue (to the first approximation) [112]. We observe that the patterns in the graphical representations of the normal mode are nearly independent of the cutoff values but the effect of cAMP binding varies slightly with the model (see Table A.1 and A.2).

Figure A.1 shows residue mobilities (square displacements of the individual residues induced by a specific mode) for the slowest mode and average mobilities for modes 1-3. It can be seen that the central helices move very little compared to the β−sheet

mode	Apo-CAP		CAP-cAMP$_1$		CAP-cAMP$_2$	
	7.3 Å	10 Å	7.3 Å	10 Å	7.3 Å	10 Å
1	0.054	0.561	0.069	0.597	0.111	0.652
2	0.154	0.752	0.189	0.781	0.209	0.808
3	0.260	1.430	0.266	1.433	0.294	1.454
4	0.275	1.555	0.290	1.573	0.299	1.600
5	0.302	1.705	0.382	1.762	0.389	1.767
6	0.378	2.107	0.385	2.202	0.456	2.261

TABLE A.1: First six eigenvalues as obtained from iGNM for two different cutoff values 7.3 Å and 10 Å.

mode	k_{cAMP_1}/k_{apo}			k_{cAMP_2}/k_{cAMP_1}		
	7.3 Å	10 Å	ElNemo	7.3 Å	10 Å	ElNemo
1	1.278	1.065	1.006	1.606	1.092	1.008
2	1.225	1.040	1.022	1.104	1.034	1.020
3	1.022	1.002	1.019	1.103	1.015	1.024
4	1.056	1.011	1.010	1.030	1.018	1.009
5	1.266	1.034	1.005	1.018	1.003	1.005
6	1.018	1.045	1.008	1.183	1.027	1.008

TABLE A.2: Ratios of the first six eigenvalues obtained from iGNM for two different cutoff values 7.3 Å and 10 Å and from ElNemo. It is evident from the iGNM results that the binding is "felt" by the model, but ElNemo appears insensitive to the binding. This is probably due to crude coarse-graining in the RTB model (see Sec. 2.3.2).
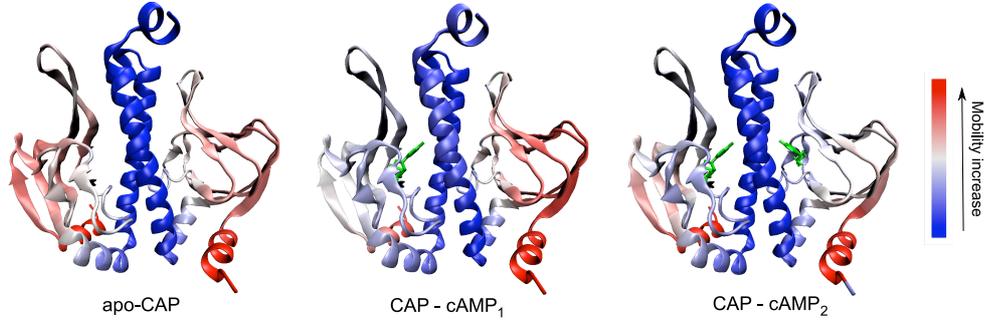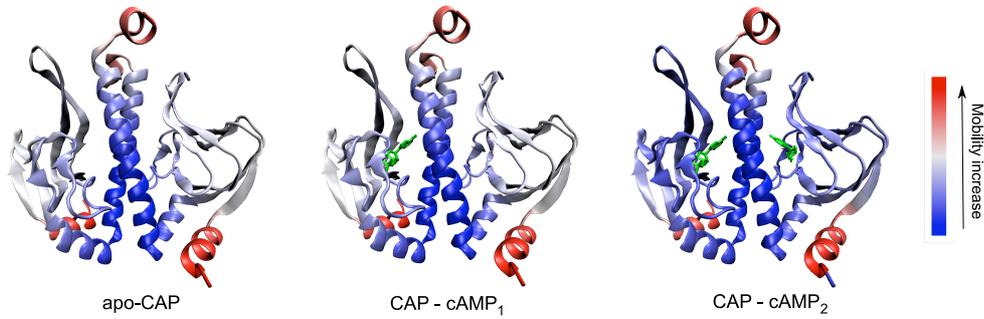
FIGURE A.1: Residue mobilities (square displacements of the individual residues) induced by a) 1st mode and b) average of 1st-3rd mode mapped onto the 3D structure of CAP^N. Low mobility is indicated by blue, high mobility by red. VMD was used for the visualisation.

complex composing the rest of the CAP monomer. The effect of ligand binding on the residue mobilities can be better observed from Fig. A.2 where the same mobilities are plotted as a function of the residue number. The first cAMP binding introduces an obvious asymmetry to the dynamic pattern of CAP. The liganded monomer becomes

(a)

(b)
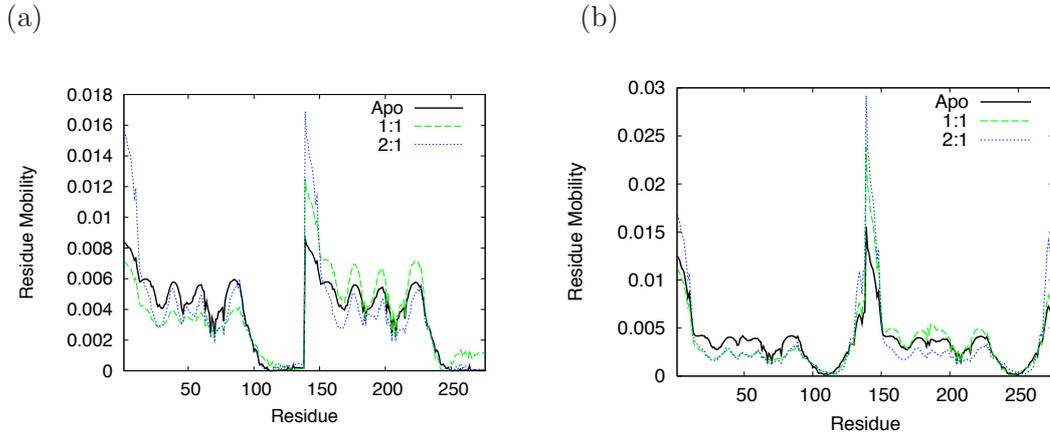
FIGURE A.2: Residue mobilities (square displacements of the individual residues) induced by (a) 1st mode and (b) average of 1st-3rd mode obtained from the iGNM server. The three ligation states of CAP: apo-CAP (in black), singly liganded CAP (in green) and doubly liganded protein (in blue) are shown. In the singly liganded protein cAMP is associated with the first monomer (residues 1-138), mobilities of cAMP itself are not shown. (a) The prominent peaks at residues 1-7 and 139-145 correspond to the small α helices at the N-terminus. A discrepancy is observed between several iGNM files as some show the fluctuations of these helices substantially smaller. Other data agree. (b) Increased mobility is observed also at the C-terminus of each chain (residues 125 -138) corresponding to the random coil at the end of the central helices.

much stiffer, the ligand provides an anchor between the previously loosely attached β structure and the central helices. The unliganded monomer's motions on the other hand become activated upon this binding. The entropic penalty for suppressing the amplified motions in the second monomer is thus higher, resulting in negative cooperativity.

Additionally, cross correlations $\Lambda_{ij} = \langle \Delta R_i \cdot \Delta R_j \rangle$ between residues i and j have been measured and are shown in Figs. A.3 and A.4. In Fig. A.3 the cross correlation map of the apo-CAP have been analysed and the strongly correlated parts of the protein have been mapped onto a 3D structure of the protein. We see that almost all correlations are confined within the individual monomers and are particularly restricted to the β-sheet structure. This implies that the two monomers are nearly uncorrelated thus moving as weakly coupled individual units. Figure A.4 shows the evaluation of dynamics upon the sequential cAMP binding. The binding introduces correlation between the central helices with the beta sheet structure of the liganded monomer. The correlation within the beta structure of the liganded monomer is disturbed yet the correlations of the unliganded monomer are unperturbed. Binding of the second cAMP re-establishes symmetry, but the correlations become weaker compared to the apo-protein.
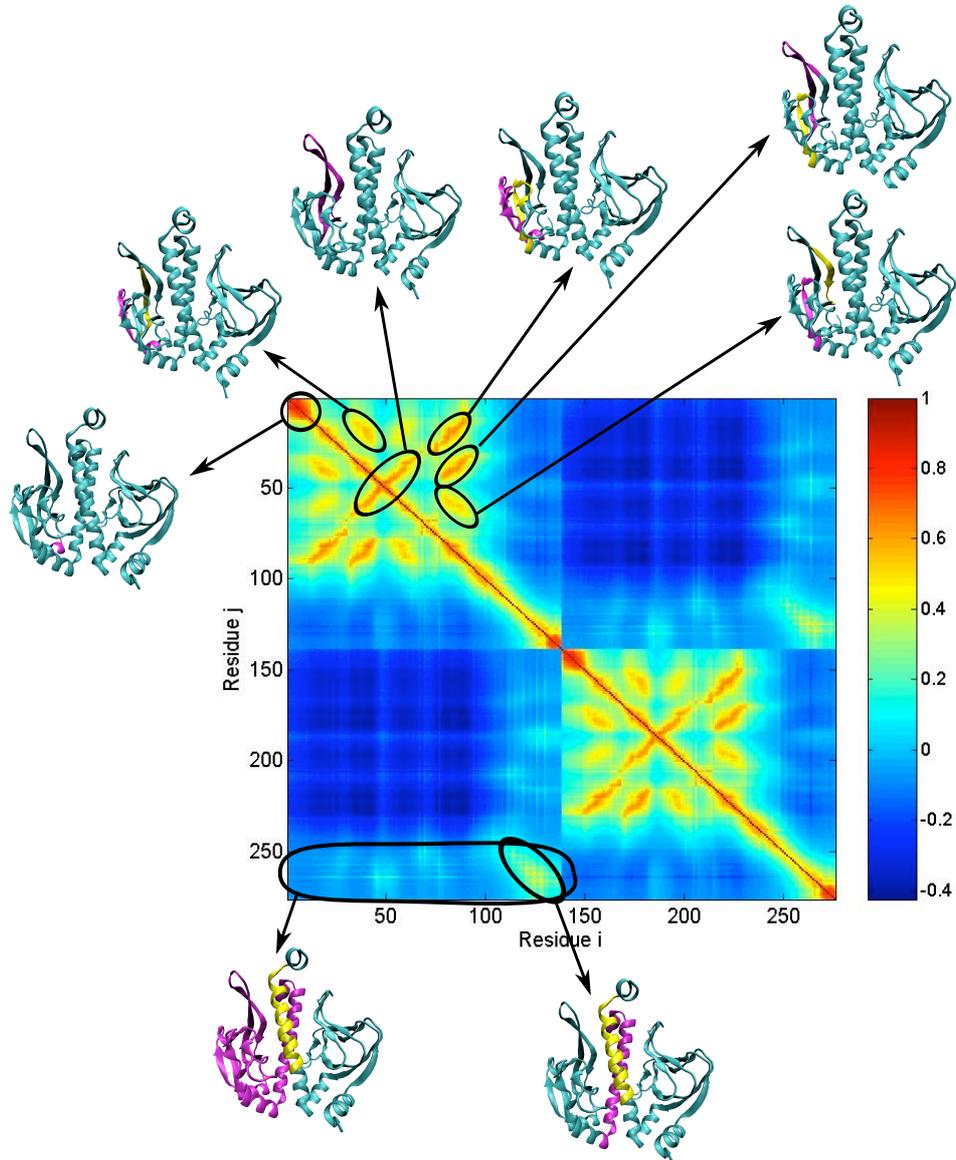
FIGURE A.3: Cross correlation map, Λ_{ij}, between residue i and j for apo-CAP$^{\text{N}}$. A pair subjected to a fully correlated motion ($\Lambda_{ij} = 1$) is colored dark red, fully anti-correlated motions ($\Lambda_{ij} = -1$) are not present and moderately correlated motions are colored dark blue. Patches of highly correlated motions are highlighted on the 3D structure of the protein in magenta and yellow. Most correlation occurs internally, inside each monomer. The two monomers are nearly uncorrelated apart from moderate correlation between the central helices.
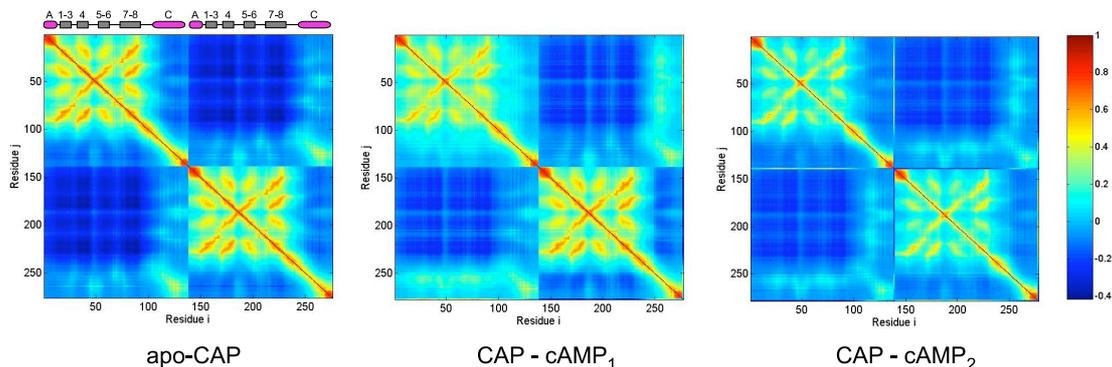
FIGURE A.4: Cross correlation map, Λ_{ij}, between residue i and j, for three ligation states of CAP^N. A pair subjected to a fully correlated motion ($\Lambda_{ij} = 1$) is colored dark red, fully anti-correlated motions ($\Lambda_{ij} = -1$) are not present and moderately correlated motions are colored dark blue. cAMP binding disturbs correlations in the liganded monomer (top left corner of the middle picture) but introduces correlation between the central helices and the liganded monomer. Binding of the second cAMP re-establishes symmetry in the motion pattern and removes correlations of the central helices to the β-sheet structures. Main parts of the secondary structure of CAP are shown above the Apo-CAP map, α-helices are represented as magenta cylinders and β-sheets as grey rectangles. A detailed sequence analysis is shown in Fig. A.9.

WEBnm and ElNemo

These two servers perform normal mode analysis (NMA) with slightly different mechanical models of the protein. WEBnm uses C_α representation of CAP^N and ElNemo uses the RTB representation (see Sec. 2.3.2), where apo-CAP^N is coarse-grained into 138 blocks containing 10-25 atoms and for the liganded versions. Each cAMP ligand is also coarse-grained into a block. The analysis performed on the individual servers is slightly different too. Eigenvalues can only be obtained from ElNemo and the ratios between the eigenvalues are compared to the iGNM values in Table A.2. Both servers output square displacements induced by the lowest modes and animations of the lowest modes. These are nearly identical at both servers and we therefore show only the results obtained from ElNemo.

The square displacements induced by the first mode are shown in Fig. A.5. The animations of the lowest modes confirm that the β–sheet structures move as semi-rigid domains with respect to the central helices. The central helix (res. 110-136, 248-274) is nearly stationary in the first mode which can be seen as a dip in Fig. A.5. Similar dip is

observed in all three lowest modes. The pattern of the mobilities is qualitatively similar to that obtained from iGNM for the unliganded protein.

Fig. A.5 reveals that the normal modes obtained from NMA are not altered by cAMP binding. This suggests that NMA as implemented on these servers is not appropriate technique for studying the allostery in CAP.
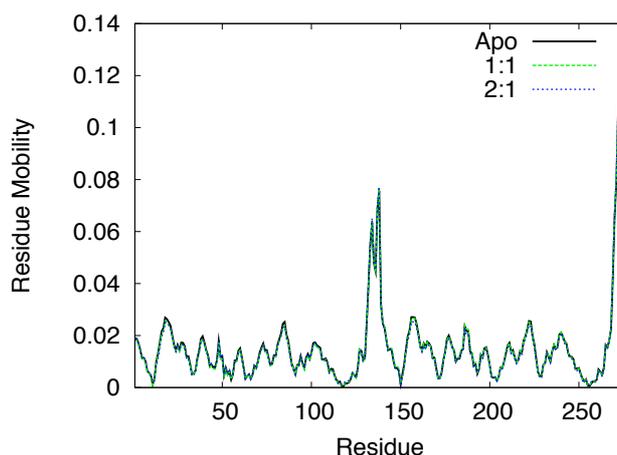
FIGURE A.5: The square displacements induced by the first mode obtained from NMA performed at the server ElNemo. The binding of the ligands is not reflected in the square displacements.

FIRST/FRODA

We performed a structure analysis of the three binding states with the program FIRST. The technique is introduced in detail in Sec. 2.4. The only variable parameter of FIRST is the energy cutoff value. The larger the value the more bonds are excluded from the calculation and therefore the identified rigid regions are smaller. Decomposition of the apo-CAP into rigid regions for cutoff values of -1.5 and -2 kcal/mol is shown in Fig. A.6. For the smaller cutoff (-1.5 kcal/mol) the central helix and the beta sheet complex are identified as one rigid structure. By increasing the cutoff, which corresponds to removing low energy bonds from the calculation each central helix becomes a the only large rigid structure. This indicates that the beta sheet structure is moving separately to the helix however is quite flexible on its own.

A graphical representation of the decomposition of all three ligation states into rigid domains for cutoff -2.5 kcal/mol is shown in Fig. A.7. In the apo-structure the two central helices move independently but in the liganded protein they become tightly bound. A part of the cAMP ligand belongs to the rigid core, whereas the other part is floppy. The decomposition of all three ligation states as a function of energy cutoff is shown in the hydrogen dilution plots in Fig. A.8.
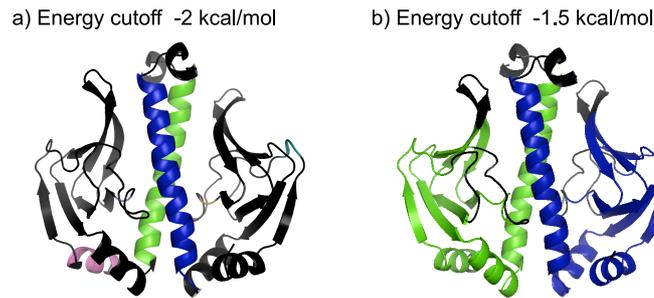
FIGURE A.6: Decomposition of apo-CAP into rigid domain performed by FIRST for two different energy cutoffs, -2 and -1.5 kcal/mol. The smaller cutoff (-1.5 kcal/mol) includes more bonds into the calculation and therefore larger regions are determined as rigid. Only regions larger than 20 atoms are shown, the rest is coloured black. The visualisation was done in PyMOL.
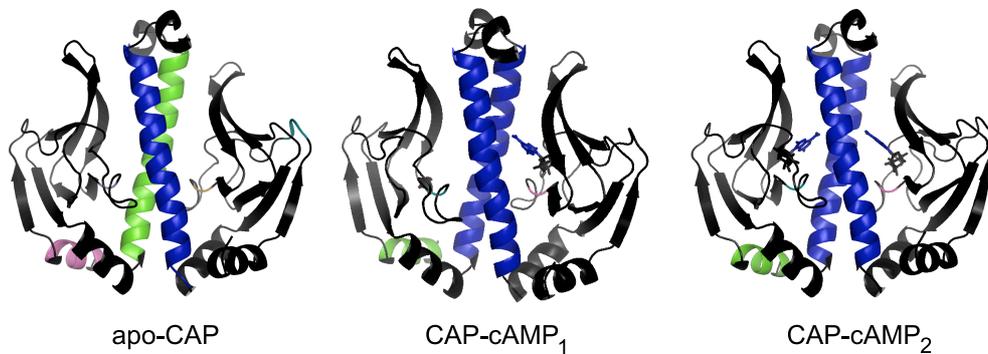
FIGURE A.7: Decomposition of the three ligation states of CAP into rigid domain performed by FIRST. The energy cutoff was -2.5 kcal/mol. Only regions larger than 20 atoms are shown, the rest is coloured black. The visualisation was done in PyMOL.

(a) apo-CAP
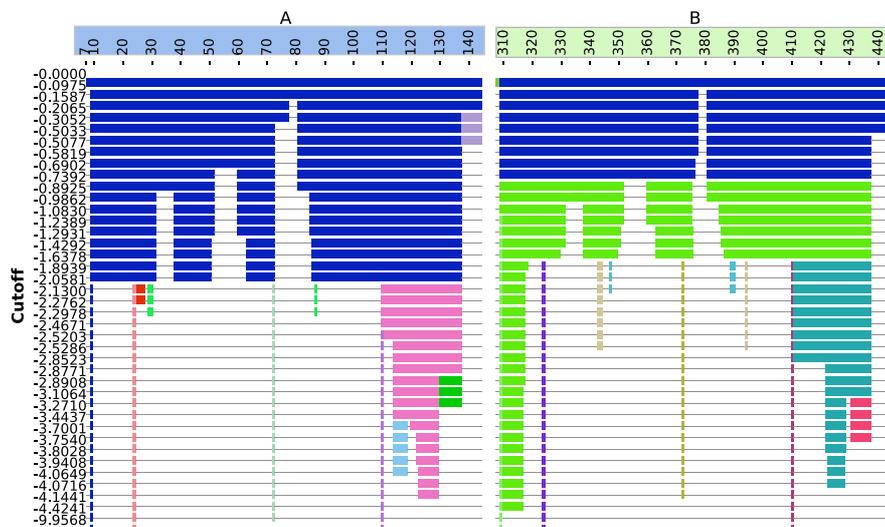
(b) CAP-cAMP$_1$
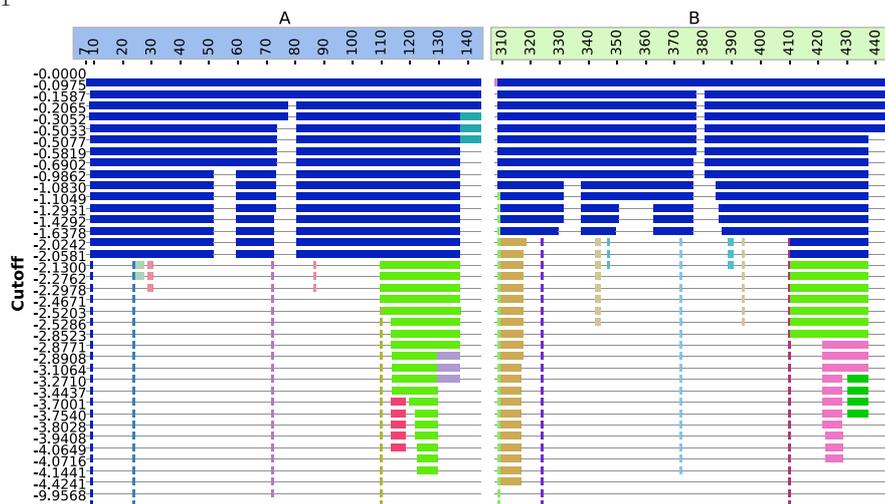
(c) CAP-cAMP$_2$
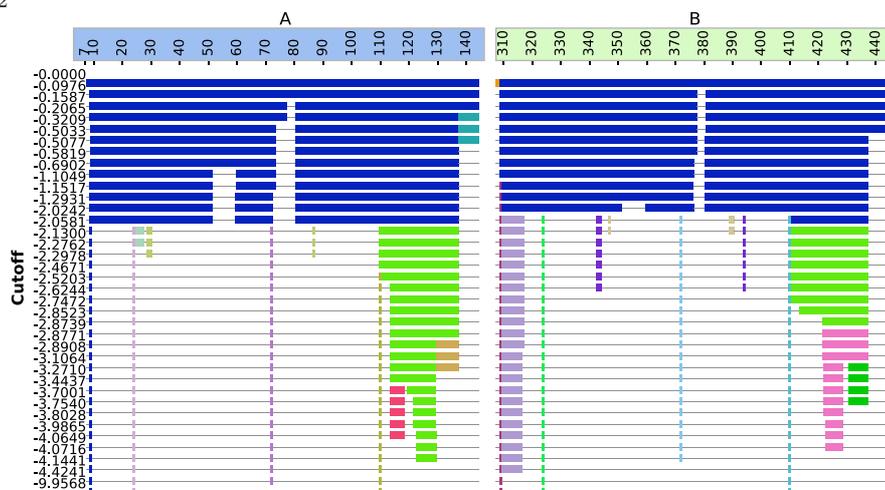
FIGURE A.8: Hydrogen dilution plots of apo-CAP (top), CAP-cAMP$_1$ (middle) and CAP-cAMP$_2$ (bottom) obtained from FIRST. The vertical axis shows the energy cutoff and the horizontal residue number. Colour bars indicate which residues belong to particular rigid domain.
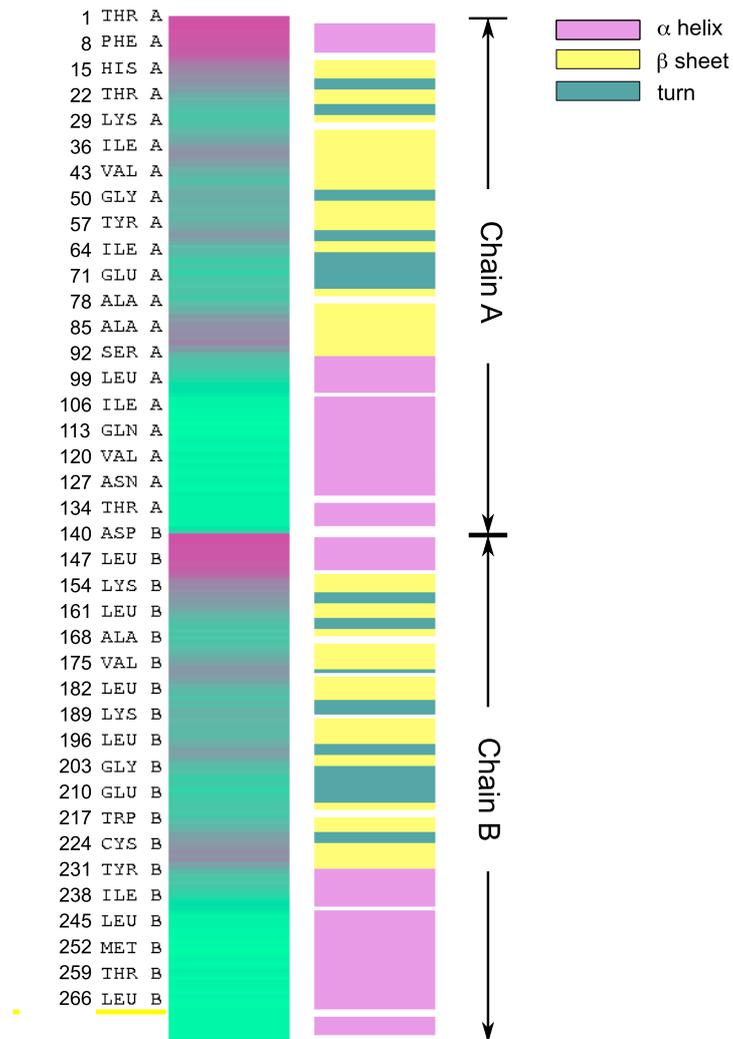
FIGURE A.9: Sequence analysis of the CAP protein obtained from VMD.

Conclusions

We established that the Gaussian network model is most suitable and beneficial model for the study of allostery in CAP. Within this model, the ligand binding is reflected in the dynamic pattern although differs slightly from the experimental observation of the group of Kalodimos [4]. In particular the loosening-tightening effect is observed only in one monomer. This may be due to the origin of the structures analysed. They were obtained by adaptation of the doubly liganded structure of the full length protein. The real structures are probably slightly different as suggested by the recent NMR structure of the apo-CAP [142].

Elastic network models indicate the same structure of the lowest frequency motions as the GNM, i.e. the β−sheet structure moving with respect to the central helices. However the ligand binding is not reflected in the dynamics and the results from ENMs are thus not considered any further.

Finally, the system was analysed by the rigid domain decomposition software FIRST. The program decomposes each monomer into two rigid blocs, the central α−helix and the β-sheet structure.

In conclusion these models confirm our hypothesis that CAP is composed of two weakly coupled monomers and each of the monomers is divided into two relatively rigid domains that move with respect to each other.

Appendix B

Elastic Theory of a Rod

In this thesis we are using several results from the theory of elasticity, a branch of physics that focuses on the mechanics of solid bodies, treating them as continuous media. A rigorous introduction into the theory can be found for example in the Landau and Lifshitzs Course of Theoretical Physics [199]. Here we derive only the few results we use in this thesis.

The fundamental concept in the theory of elasticity is a deformation. A solid body becomes deformed under an action of applied force, this deformation generally includes a change in shape and volume. In order to mathematically describe the deformation let us consider any point in the body. Before the deformation this point is defined by a position vector $\mathbf{r}$ and after the deformation by a vector $\mathbf{r}'$. Let us introduce the displacement vector $\mathbf{u}$ as

$$\mathbf{u} = \mathbf{r}' - \mathbf{r} \tag{B.1}$$

It proves useful to define a so called strain tensor u_{ik} that describes a deformation in a small length element of the body. Using the Einstein summation convention this tensor is

$$u_{ik} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} + \frac{\partial u_l}{\partial x_i} . \frac{\partial u_l}{\partial x_k} \right) \tag{B.2}$$

Evidently the strain tensor is symmetric $u_{ik} = u_{ki}$. In this thesis we deal only with small deformations, for which the change of the deformation vector $\partial u_i / \partial x_k$ is also small and we can therefore neglect the last term in Eq. (B.2). The tensor then reduces to

$$u_{ik} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i} \right) \tag{B.3}$$

199

When a deformation occurs, the internal structure of the body is changed and forces arise that tend to return the body back to equilibrium. It can be shown [199] that the force per unit volume F can be expressed as a divergence of a tensor of rank two

$$F_i = \frac{\partial \sigma_{ik}}{\partial x_k}, \tag{B.4}$$

where the Einstein summation rule is used. The tensor σ_{ik} is usually referred to as the stress tensor. It can be seen that the component σ_{ik} of the stress tensor is the i−th component of the force on unit area perpendicular to the k axis.

The stress tensor can also be obtained by differentiating the Helmholtz free energy $\mathcal{F}$ at constant temperature

$$\sigma_{ik} = \left(\frac{\partial \mathcal{F}}{\partial u_{ik}} \right)_T . \tag{B.5}$$

We now wish to directly relate the strain and stress tensor. The deformation is small and we can therefore expand the free energy in powers of u_{ik} around 0. The linear term vanishes because $\partial \mathcal{F}/\partial u_{ik} = \sigma_{ik}$ and $\sigma_{ik} = 0$ when $u_{ik} = 0$ (undeformed body). Including all the allowed second order terms the free energy becomes

$$\mathcal{F} = \mathcal{F}_0 + \frac{1}{2}\lambda u_{ii}^2 + \mu u_{ik}^2, \tag{B.6}$$

λ and μ are called the Lamé coefficients. Using the equation (B.5) and some tensor algebra we find that

$$\sigma_{ik} = K u_{ll}\delta_{ik} + 2\mu\left(u_{ik} - \frac{1}{3}\delta_{ik}u_{ll}\right). \tag{B.7}$$

δ_{ik} is the Kronecker delta and the coefficient $K = \lambda + 2/3\mu$ is called the bulk modulus. This expression can be inverted resulting in

$$u_{ik} = \frac{1}{9K}\delta_{ik}\sigma_{ll} + \frac{1}{2\mu}\left(\sigma_{ik} - \frac{1}{3}\delta_{ik}\sigma_{ll}\right). \tag{B.8}$$

We see that the strain tensor u_{ik} is a linear function of the stress tensor σ_{ik}. This is equivalent to saying that the deformation is proportional to the applied forces. This conclusion, valid for small deformations, is the general formulation of the Hooke's law.

Another useful formula follows from the fact that $\mathcal{F}$ is a quadratic function of u_{ik} and therefore $u_{ik}(\partial \mathcal{F}/\partial u_{ik}) = 2\mathcal{F}$. Using the equation (B.5) we arrive at

$$\mathcal{F} = \frac{1}{2}\sigma_{ik}u_{ik} \tag{B.9}$$
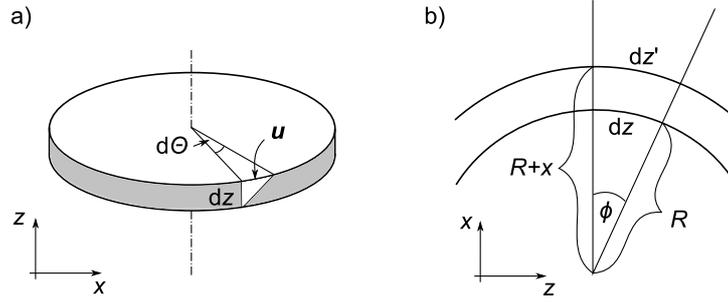
FIGURE B.1: a) Incremental torsion deformation, b) incremental bending deformation.

Now let us use the derived equations to find the energy of deformation of thin rods. We are interested in two types of deformations, the twisting and the bending. We start be deriving the energy of a twisted rod.

Twisting of a Thin Rod

Here we derive the energy of torsional deformation. By torsionally deformed rod we mean a straight rod with each transverse section rotated with respect to the next section. We assume a coordinate system with the z-axis pointing along the axis of the rod. A small section of the rod is shown in Fig. B.1 a. In this layer of thickness dz the top is rotated relative to the bottom by a fractional angle $d\Theta$. Let us introduce the longitudinal density of twisting $\tau = d\Theta/dz$. The radius vector $\mathbf{r}$ is rotated through the angle $d\Theta$ resulting in the displacement vector

$$\mathbf{u} = \mathbf{d\Theta} \times \mathbf{r}, \tag{B.10}$$

where $\mathbf{d\Theta}$ is a vector with magnitude $d\Theta$ and direction of the axis of rotation; in our case $\mathbf{d\Theta} = (0, 0, \tau z)$. Carrying out the cross product we find the components of the displacement vector

$$\mathbf{u} = (-\tau y z, \tau z x, 0), \tag{B.11}$$

where we assumed that the points in the rod do not undergo any displacement along the z-axis. Generally this is not true, however the expression for the energy of the twisted rod remains unchanged even when displacements along the z-axis takes place [199].

From the displacement vector we derive the strain tensor by taking the appropriate derivatives (Eq. (B.3)). The only non-zero elements are

$$u_{xz} = \frac{1}{2}\left(\frac{\partial u_x}{\partial z} + \frac{\partial u_z}{x}\right) = -\frac{1}{2}\tau y,$$
$$u_{yz} = \frac{1}{2}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{y}\right) = \frac{1}{2}\tau x. \tag{B.12}$$

In order to calculate the free energy of the deformation we need to find the stress tensor. Using the inverted Hooke's law Eq. (B.7)

$$\sigma_{zy} = 2\mu u_{zy} = \mu\tau x, \sigma_{xz} = -\mu\tau y.$$

Eq. (B.9) gives the free energy per unit volume as

$$\mathcal{F} = \frac{1}{2}\sigma_{ik}u_{ik} = \frac{1}{2}\mu\tau^2(x^2 + y^2). \tag{B.13}$$

The free energy per unit length is obtained by integrating over the surface of the cross-section S'

$$\mathcal{F} = \int_{S'} \frac{1}{2}\mu\tau^2(x^2 + y^2)\mathrm{d}S \equiv \frac{1}{2}\mu J\tau^2, \tag{B.14}$$

where J is the polar moment of inertia. For a circular cross section of radius R the moment of inertia $J = \pi R^4/2$. The twist density τ in principle varies along the z-axis, therefore the total free energy of torsionally deformed rod is

$$\mathcal{F}_{rod} = \frac{1}{2}\int_0^L C\tau^2\mathrm{d}z, \tag{B.15}$$

where we introduced the twisting rigidity $C = \mu J$.

Bending of a Thin Rod

Now let us consider a thin rod that is slightly bent. We introduce a coordinate system such that the z-axis runs along the axis of the rod when straight and the rod is bend the zx-plane. The bending introduces compression in some parts of the rod and extension in others. We can locate a neutral surface somewhere midway through the rod where neither extension nor compression occurs. All points on the concave side bellow the neutral surface are compressed and all points on the convex side are extended.

Moreover we assume that all internal stresses occur only in the z direction and therefore the only non-zero element of the stress tensor is σ_{zz}. This seems like a reasonable assumption, but it can be shown more rigorously. The rod is thin and consequently the external forces applied to the sides of it are much smaller than internal forces inside the rod and can be taken as zero. In general, let $\mathbf{P}$ be an external force on the unit area of surface of the body. The force $\mathbf{P}dS$ acting on the surface element dS has to be balanced by the internal force $-\sigma_{ik}dS_k$. Let $\mathbf{n}$ be the normal vector to the surface element dS, then $dS_k = n_k dS$ and the force balance results in the boundary condition

$$\sigma_{ik}n_k = P_i. \tag{B.16}$$

In our case then $\sigma_{ik}n_k = 0$ everywhere on the sides of the rod and $n_z = 0$. Therefore $\sigma_{xx}n_x + \sigma_{xy}n_y = 0$ and similarly for $i = y, z$. Let us consider a point on the surface where the normal is parallel to the x-axis. At this point also $n_y = 0$ and the above equation yields $\sigma_{xx} = 0$. There will be two such points on the opposite sides of the rod. Since the rod is thin and the deformation small, σ_{xx} has to be small everywhere inside the rod too and we can therefore neglect it and set $\sigma_{xx} = 0$ everywhere in the rod. Similarly we can find that all components of the stress tensor apart from σ_{zz} are equal zero.

In order to determine the free energy of the bend rod we need to find the u_{zz} element of the strain tensor, which is the relative extension along the z-axis. Let us consider an element of length dz. Upon bending this element is deformed into dz'. The only elements that remain unchanged are those at the neutral surface. Let R be the radius of curvature of the neutral surface, the dz and dz' can be seen as arc lengths of radii R and $R + x$ respectively (see Fig. B.1 b). The relative extension is then found from comparing the tangents $\tan\phi = dz/R = dz'/(R + x)$ which yields

$$u_{zz} = \frac{dz' - dz}{dz} = \frac{x}{R}. \tag{B.17}$$

Using the Hooke's law Eq. (B.7) we find that

$$\sigma_{zz} = E u_{zz}, \tag{B.18}$$

where we used the proportionality constant E usually referred to as the Young's modulus. The bending free energy per unit volume is then

$$\mathcal{F} = \frac{1}{2}\sigma_{zz}u_{zz} = \frac{1}{2}E\frac{x^2}{R^2} \tag{B.19}$$

In order to obtain the free energy per unit length we integrate over the cross section. Integrating over the cross section

$$\frac{1}{2}\frac{E}{R^2}\int_{S'} x^2 \mathrm{d}S \equiv \frac{1}{2}\frac{E}{R^2}J_y, \tag{B.20}$$

where J_y the moment of inertia about the y-axis. For a circular cross section $J_y = \pi R^4/4$ and the total free energy of bend rod is given by

$$\mathcal{F}_{rod} = \frac{1}{2}\int_0^L \frac{A}{R^2}\mathrm{d}z \tag{B.21}$$

where the bending rigidity $A = EJ_y$ has been introduced.

Appendix C

Derivation of Bending Eigenmodes

In a thermal environment the local curvature of a rod constantly fluctuates. The shape of the rod at any time is given as a linear combination of the normal modes. Finding the normal modes corresponds to determining eigenfunctions of the Hamiltonian E_b (Eq. (4.2)). The term eigenfunctions of the Hamiltonian is imprecise, rigorously we mean finding eigenfunctions of the differential operator $\mathcal{L}$ defined by the equation

$$E_b = \int_0^L h\,\mathcal{L}\,h\,\mathrm{d}x. \tag{C.1}$$

In order to convert Eq. (4.2) into this form we integrate it twice by parts.

We then search for solutions of the eigenvalue problem of $\mathcal{L}$ which reads

$$\frac{1}{2}Ah^{(4)}(x) = \lambda h(x) \tag{C.2}$$

The general solution is a linear combination of four functions

$$h(x) = C_1\cos(\sigma x) + C_2\sin(\sigma x) + C_3\cosh(\sigma x) + C_4\sinh(\sigma x), \tag{C.3}$$

where $\sigma = \sqrt[4]{2\lambda/A}$.

We determine σ and C_i by applying the boundary conditions and normalization of the eigenfunctions. The free ends boundary conditions read

$$\left.\frac{\partial^2 h}{\partial x^2}\right|_{x=0,L} = 0$$

$$\left.\frac{\partial^3 h}{\partial x^3}\right|_{x=0,L} = 0 \tag{C.4}$$

The boundary conditions say that both, the bending moment and the shear force have to vanish at the ends. This yields

$$h_n(x) = (\cos(\sigma_n L) - \cosh(\sigma_n L))(\cos(\sigma_n x) + \cosh(\sigma_n x))$$
$$+ (\sin(\sigma_n L) + \sinh(\sigma_n L))(\sin(\sigma_n x) + \sinh(\sigma_n x)), \tag{C.5}$$

where σ_n is the nth root of the equation

$$\cosh(\sigma L)\cos(\sigma L) = 1. \tag{C.6}$$

This equation can only be solved numerically and the first few solutions are

$$\sigma_n L \doteq 4.73, 7.853, 10.996, \ldots \tag{C.7}$$

For higher n the solutions can be approximated by a simple formula

$$\sigma_n L \doteq \frac{(2n+1)\pi}{2} \tag{C.8}$$

This approximation is very good yielding the first few values 4.71, 7.854 and 10.996. The two sequences converge with increasing n. However the function h_n becomes very hard to manipulate with increasing n, because of the exponentially diverging hyperbolic functions. The functions h_n can however be approximated for large n. We normalize h_n by its value at $x = 0$ and from the condition $\cos(\sigma_n L)\cosh(\sigma_n L) = 1$ obtain

$$\cos(\sigma_n L) - \cosh(\sigma_n L) = -\frac{\sinh^2(\sigma_n L)}{\cosh(\sigma_n L)}$$

$$\sin(\sigma_n L) + \sinh(\sigma_n L) = \frac{\sinh(\sigma_n L)}{\cosh(\sigma_n L)}(1 + \cosh(\sigma_n L)) \tag{C.9}$$

Inserting these expressions back into Eq. (C.5) yields

$$
h_n = \frac{1}{2\sinh(\sigma_n L)} \left[(\cos(\sigma_n x) + \cosh(\sigma_n x)) \sinh(\sigma_n L) \right.
$$
$$
\left. - (1 + \cosh(\sigma_n L)) (\sin(\sigma_n x) + \sinh(\sigma_n x)) \right] \tag{C.10}
$$

We rewrite the hyperbolic functions using exponential functions and rearrange to yield

$$
\sinh(\sigma_n L) = \frac{1}{2} e^{\sigma_n L} (1 - \varepsilon)
$$
$$
\sinh(\sigma_n x) = \frac{1}{2} e^{\sigma_n x} (1 - \tilde{\varepsilon}), \tag{C.11}
$$

and similarly for the hyperbolic cosine. In the equation we introduced factors $\varepsilon = e^{-2\sigma_n L}$, $\tilde{\varepsilon} = e^{-2\sigma_n x}$ which tend to 0 for $n \to \infty$. By inserting these into Eq. (C.10) and simplifying the expression we obtain the approximated functions h_n

$$
h_n(x) = \frac{1}{2} (\cos(\sigma_n x) - \sin(\sigma_n x)) \tag{C.12}
$$

The comparison of the original and approximated expressions are shown in Fig. 4.6.

Similarly the eigenfunctions of the twisting energy E_t are found by solving the eigenvalue problem

$$
C\theta'' = \lambda\theta, \tag{C.13}
$$

Again, we use the free ends boundary conditions

$$
\left. \frac{\partial \theta}{\partial x} \right|_{x=0,L} = 0 \tag{C.14}
$$

to obtain

$$
\theta_n = B_n \cos\left(\frac{n\pi x}{L}\right), \qquad n = 1, 2, \ldots \tag{C.15}
$$

Appendix D

Analysis of Eigenvalues of the Twisting Modes

We expressed the Hamiltonians of twisting motion of the elastic rod in the singly and doubly liganded state in the orthogonal basis of the eigenvectors of the homogeneous rod. In order to apply the calculation to real molecules we needed to find the first N eigenvalues of the infinite matrix, or in other words, the first N true eigenvalues of the singly and doubly bound Hamiltonians. We proceeded by truncating the matrix, calculating the eigenvalues of this matrix and comparing them to the true eigenvalues. We found that only the high eigenvalues are affected by the truncation and are separated from the correct eigenvalues by a sharp "kink" (see Fig. 4.7).

In a given geometry, the "kink" consistently occurs at a certain position of the rank of the matrix. This can be seen from the plot of the kink location as a function of the rank of the truncated matrix (Fig. D.1). For an example set of parameters ($a = 1/3$, $d = 1/6$ and $\gamma = 20$), we find a linear dependence with a slope of 0.7. The N-dependence is found to be linear for all values of the parameters but with a slightly varying slope. The kink location is also γ-dependent, this dependence can be fitted with an inverse exponential function as shown in Fig. D.2.

From Fig. D.2 we also observe that the "kink" occurs at much lower mode number in the doubly liganded rod. When evaluating the allosteric free energies the number of modes included must be the same in all states. In practice we choose the dimension of the truncated matrix large enough to obtain a sufficient amount of correct eigenvalues.
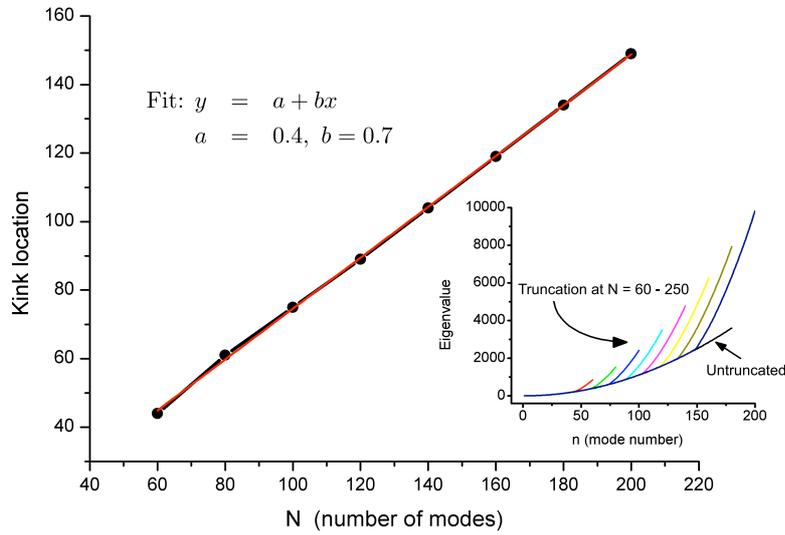
FIGURE D.1: The "kink" location as a function of number of modes (dimension of the truncated matrix) N in black. The dependence is linear (red fit line), the fit parameters are presented in the table in the top left corner. The kinks are found from the inset graph (bottom right).
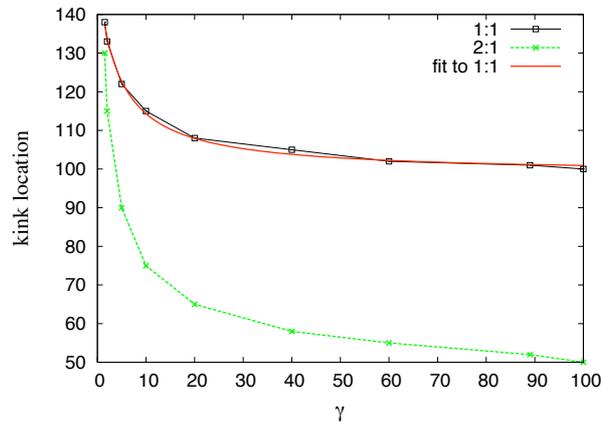
FIGURE D.2: The "kink" location for the singly (1:1) and doubly (2:1) liganded elastic rod as a function of parameter γ. Both functions can be fitted with a function $f(x) = a + \frac{b}{\exp(c\gamma)+d}$ as shown for the case 1:1 (red line). The fit parameters were $a = 99$, $b = 0.56$, $c = 0.003$ and $d = -1$. The rank of the truncated matrix was 150.

Appendix E

Evaluation of Elastic Parameters of the Groove Breathing Model

The elastic parameters (ξ/L, α and β) of the groove breathing model were extracted from the 11 ns simulation of Harris *et al.* [101] using the following method.

The energy of the fluctuating groove width f was derived in Sec. 5.2.1 and yields

$$E_{gw} = \int\limits_0^1 \left[\frac{k}{2} f'(x)^2 + \lambda f(x)^2 \right] \mathrm{d}x \qquad \text{(E.1)}$$

The groove width f can be written as a sum of N Fourier modes

$$f(x) = \sum_{n=0}^{N} a_n \cos(2\pi n x), \qquad \text{(E.2)}$$

and the amplitudes are found from

$$a_n = 2 \sum_{i=1}^{N} f(x_i) \cos(2\pi n x_i), \qquad \text{(E.3)}$$

where x_i is the location of the i-th measurement point. We can only measure the groove width at nine points (there are 12 base pairs in the oligomer) and thus have to use $N = 9$. The groove width was measured as a distance between i-th and $(i+3)$-rd phosphorus on the opposite backbone (Fig. 5.1). The locations x_i of these 9 measurement points were

found by projecting the middle of the distance between i-th and $(i + 3)$-rd phosphorus onto the helical axis.

The energy in terms of the Fourier modes is calculated by inserting (E.2) into the energy expression (E.1) and assuming that k and λ are constant along the whole DNA helix. The equipartition theorem states that there is energy of $k_B T/2$ available for each mode, which leads to

$$\left\langle a_n^2 \right\rangle = \frac{k_B T}{\frac{\lambda}{2} + 4k\pi^2 n^2}. \tag{E.4}$$

We extract the average value of a_n^2 by inserting the measured widths into (E.3) and averaging over the course of the simulation. The values of $\left\langle a_n^2 \right\rangle$ have been found for various window widths and statistical errors calculated as standard deviations from these values. Each binding state was simulated for ~ 11 ns corresponding to 11000 frames. The first 1000 frames were disregarded and the rest was divided into windows of 3000 - 5000 frames. In order to capture the crucial slow modes, we did not reduce the window width any further. The resulting $\left\langle a_n^2 \right\rangle$ values and their corresponding errors change negligibly with the window width (data not shown).

The $\left\langle a_n^2 \right\rangle$ dependence on n was fitted to Eq. (E.4). The best fit implies that $\xi/L = 0.06$, but reasonable fits are obtained for ξ/L ranging from 0.06 to 1.5. The fits are shown in Fig. E.1 a. We repeat the process for the 2:1 complex and again assume that the elastic
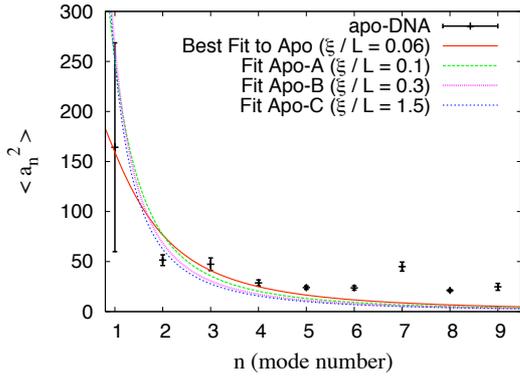
a) apo-DNA   b) 2:1 complex
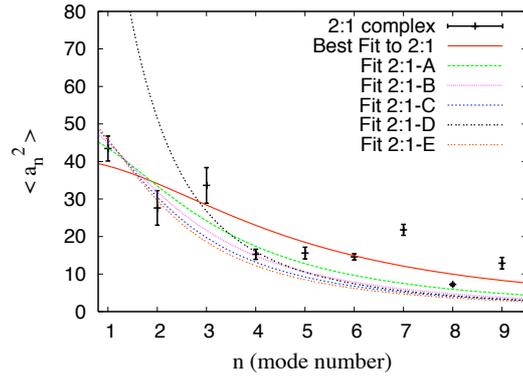
FIGURE E.1: The averaged Fourier coefficients $\left\langle a_n^2 \right\rangle$ of the minor groove width measured from the simulation of DNA-Hoechst complexes. The values of $\left\langle a_n^2 \right\rangle$ of a) the apo-DNA and b) the 2:1 complex were obtained as an average value from window widths of 3000 frames. For the details about the fits see main text and Table E.1.

| Apo fit	2:1 fit	ξ/L	α	β	$\Delta\Delta G/k_B T$
best	best	0.06	0.5	6.3	3.1
best	A	0.06	1	5.3	0.0003
A	best	0.1	0.4	25	5.5
A	B	0.1	1	19	-0.03
B	best	0.3	0.3	12	10.1
B	C	0.3	1	90	-0.2
B	D	0.3	1	25	-0.14
C	best	1.5	0.3	2500	8.6
C	E	1.5	1	1800	-1.2

TABLE E.1: Model parameters obtained from fitting Fourier coefficients of the apo and 2:1 complex of the oligomer groove widths. The fits are shown in Fig. E.1. The geometric parameters used to calculate $\Delta\Delta G$ were $a = 0.435$, $d = 0.13$ and $N = 70$.

parameters are constant along the whole groove. The amplitudes $\langle a_n^2 \rangle$ are functions of the multiplicative factors α and β

$$\langle a_n^2 \rangle = \frac{k_B T}{\frac{\beta\lambda}{2} + 4\alpha k\pi^2 n^2}. \tag{E.5}$$

We use the values obtained from the apo-DNA fit to extract the values of α and β. The best fits are found for $\xi/L = 0.06$, $\alpha = 0.5$ and $\beta = 6.3$.

Because of the unsatisfactory quality of the fits we estimate α from a complementary analysis. We measure the fluctuations of the atoms along the groove edges in the three binding states and find they remain constant during the binding events (Fig. 5.6 b). We conclude that it is reasonable to set the phonon stiffening factor $\alpha = 1$. Finding that the phonons remain unperturbed ($\alpha = 1$) implies that groove breathing contributes to positive allostery.

We then take different ξ/L values from the apo-DNA fits (Fig. E.1 a) and use them to construct the fits of the 2:1 structure and evaluate β under the condition that $\alpha = 1$. The data for the 2:1 amplitudes are very scattered and it is therefore difficult to quantify the quality of the fits. The fits get worse with increasing ξ/L but remain acceptable even for $\xi/L = 1.5$. The fits are presented in Fig. E.1 b. The resulting parameter values and the allosteric free energies are shown in Table. E.1.

We observe that while the fits with the fixed value of $\alpha = 1$ look acceptable, the derived values of β are very large, in particular for large ξ/L. Reduction of the value of β leads to

substantially decreased accuracy of the fit (fit D in Fig. E.1). The most probable source of systematic error in this method is the inhomogeneity of the parameters along the DNA and in the case of doubly bound oligomer the gap between ligands, which is unaccounted for in the fitting function. It is in principle possible to repeat the calculation for $\langle a_n^2 \rangle$ for the real geometry, but the measured data are so scattered and the fits so unreliable that we omit this step.

Bibliography

[1] Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H. and Phillips, D.C. 3-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature*, 181:662–666, 1958.

[2] Weber, G. Ligand binding and internal equilibria in proteins. *Biochemistry*, 11:864–878, 1972.

[3] Tokuriki, N. and Tawfik, D.S. Protein dynamism and evolvability. *Science*, 324(5924):203–207, 2009.

[4] Popovych, N., Sun, S., Ebright, R.H. and Kalodimos, C.G. Dynamically driven protein allostery. *Nat. Struct. Mol. Biol.*, 13(9):831 – 838, 2006.

[5] Jarymowycz, V.A. and Stone, M.J. Remote changes in the dynamics of the phosphotyrosine-binding domain of insulin receptor substrate-1 induced by phosphopeptide binding. *Biochemistry*, 47(50):13371–13382, 2008.

[6] Dunker, A.K., Silman, I., Uversky, V.N. and Sussman, J.L. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.*, 18(6):756–764, 2008.

[7] Smock, R.G. and Gierasch, L.M. Sending signals dynamically. *Science*, 324:198–203, 2009.

[8] Frederick, K.K., Marlow, M.S., Valentine, K.G. and Wand, A.J. Conformational entropy in molecular recognition by proteins. *Nature*, 448:325–U3, 2007.

[9] Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Korzhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E. and Kern, D. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121, 2005.

[10] Korzhnev, D.M., Salvatella, X., Vendruscolo, M., Di Nardo, A.A., Davidson, A.R., Dobson, C.M. and Kay, L.E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature*, 430(6999):586–590, 2004.

[11] Keskin, O., Bahar, I., Flatow, D., Covell, D.G. and Jernigan, R.L. Molecular mechanisms of chaperonin GroEL-GroES function. *Biochemistry*, 41(2):491–501, 2002.

[12] Noskov, S.Y. and Roux, B. Ion selectivity in potassium channels. *Biophys.Chem.*, 124(3):279–291, 2006.

[13] Diez, M., Zimmermann, B., Borsch, M., Konig, M., Schweinberger, E., Steigmiller, S., Reuter, R., Felekyan, S., Kudryavtsev, V., Seidel, C.A.M. and Graber, P. Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. *Nat. Struct. Mol. Biol.*, 11(2):135–141, 2004.

[14] Freire, E. The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. USA*, 96:10118–10122, 1999.

[15] Kern, D. and Zuiderweg, E.R.P. The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, 13:748–757, 2003.

[16] Lockless, S.W. and Ranganatha, R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science*, 286:295–299, 1999.

[17] Goodey, N.M. and Benkovic, S.J. Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.*, 4(8):474–482, 2008.

[18] Cooper, A. and Dryden, D.T.F. Allostery without conformational change - a plausible model. *Eur. Biophys. J.*, 11:103–109, 1984.

[19] Bahar, I., Atilgan, A.R., Demirel, M.C. and Erman, B. Vibrational dynamics of folded proteins: Significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.*, 80(12):2733–2736, 1998.

[20] Chaires, J.B. Allostery: DNA does it too. *ACS Chem. Biol.*, 3(4):207–209, 2008.

[21] Lesne, A. and Victor, J.M. Chromatin fiber functional organization: Some plausible models. *Eur. Phys. J. E*, 19:279–290, 2006.

[22] Creighton, T.E. *Proteins, structures and molecular properties*. W.H. Freeman and Co., 1993.

[23] Fischer, E. Einfluss der configuration auf die wirkung derenzyme. *Berichte der Deutschen Chemischen Gesellschaft*, 27:2985–2993, 1894.

[24] Koshland, D.E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. USA*, 44:98–104, 1958.

[25] Cornish-Bowden, A. Enthalpy-entropy compensation: a phantom phenomenon. *J.Biosci.*, 27:121–126, 2002.

[26] Bohr, C. and Hasselbach, K.A. Uber einen in biologischen beziehung wichtigen Einfluss, den die Kohlen-sauerspannung des Blutes auf dessen Sauerstoffbindung ubt. *Skand. Arch. Physiol.*, 15:401–412, 1904.

[27] Eaton, W.A., Henry, E.R., Hofrichter, J. and Mozzarelli, A. Is cooperative oxygen binding by hemoglobin really understood? *Nat. Struct. Biol.*, 6:351 – 358, 1999.

[28] Voet, D., Voet, J.G. and Pratt, C.W. *Fundamentals of biochemistry: life at the molecular level*. John Wiley and Sons, 2006.

[29] Adair, G.S. A critical study of the direct method of measuring the osmotic pressure of haemoglobin. *Proc. R. Soc. A*, 108A:627–637, 1925.

[30] Monod, J., Wyman, J. and Changeux, J.P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.*, 12:88–118, 1965.

[31] Koshland, D.E., Nemethy, G. and Filmer, D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*, 5:365–385, 1966.

[32] Akke, M., Bruschweiler, R. and Palmer, A.G. NMR order parameters and free-energy - an analytical approach and its application to cooperative Ca2+ binding by calbindin-d(9k). *J. Am. Chem. Soc.*, 115:9832–9833, 1993.

[33] Malmendal, A., Evenas, J., Forsen, S. and Akke, M. Structural dynamics in the C-terminal domain of calmodulin at low calcium levels. *J. Mol. Biol.*, 293:883–899, 1999.

[34] Berger, C., Weber-Bornhauser, S., Eggenberger, J., Hanes, J., Pluckthun, A. and Bosshard, H.R. Antigen recognition by conformational selection. *FEBS Lett*, 450(1-2):149–153, 1999.

[35] Volkman, B.F., Lipson, D., Wemmer, D.E. and Kern, D. Two-state allosteric behavior in a single domain signalling protein. *Science*, 291:2429–2433, 2001.

[36] Tsai, C.J., Kumar, S., Ma, B.Y. and Nussinov, R. Folding funnels, binding funnels, and protein function. *Prot. Sci.*, 8(6):1181–1190, 1999.

[37] Pan, H., Lee, J.C. and Hilser, V.J. Binding sites in Escherichia coli dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. USA*, 97:12020–12025, 2000.

[38] Hawkins, R.J. and McLeish, T.C.B. Coarse-grained model of entropic allostery. *Phys. Rev. Lett.*, 93:098104–1 – 098104–4, 2004.

[39] Gunasekaran, K.Y., Ma, B. and Nussinov, R. Is allostery an intrinsic property of all dynamic proteins? *Proteins Struct. Funct. Bioinf.*, 57:433–443, 2004.

[40] Tsai, C.J., del Sol, A. and Nussinov, R. Allostery: Absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.*, 378:1–11, 2008.

[41] Hawkins, R.J. and McLeish, T.C.B. Dynamical allostery of protein alpha helical coiled-coils. *J.R. Soc. Interface*, pages 1–13, 2005.

[42] Hawkins, R. and McLeish, T.C.B. Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophys. J.*, 91:2055 – 2062, 2006.

[43] Fastrez, J. Engineering allostering regulation into biological catalysts. *ChemBioChem*, 10(18):2824–2935, 2009.

[44] Laine, E., Goncalves, C., Karst, J.C., Lesnard, A., Rault, S., Tang, W.J., Malliavin, T.E., Ladant, D. and Blondel, A. Use of allostery to identify inhibitors of calmodulin-induced activation of Bacillus anthracis edema factor. *Proc. Natl. Acad. Sci. USA*, 107(25):11277–11282, 2010.

[45] Austin, R.H., Beeson, K.W., Eisenstein, L., Frauenfelder, H. and Gunsalus, I.C. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14:5355–5373, 1975.

[46] Lipari, G. and Szabo, A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. theory and range of validity. *J. Am. Chem. Soc.*, 104:4546–4559, 1982.

[47] Lipari, G. and Szabo, A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. analysis of experimental results. *J. Am. Chem. Soc.*, 104:4559–4570, 1982.

[48] Homans, S.W. Probing the binding entropy of ligand-protein interactions by NMR. *ChemBioChem*, 6(1585-1591), 2005.

[49] Henzler-Wildman, K. and Kern, D. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007.

[50] Daniel, R.M., Dunn, R.V., Finney, J.L. and Smith, J.C. The role of dynamics in enzyme activity. *Annu. Rev. Biophys. Biomol. Struct.*, 32:69–92, 2002.

[51] Wand, A. On the dynamic origins of allosteric activation. *Science*, 293:1395a, 2001.

[52] Cui, Q. and Karplus, M. Allostery and cooperativity revisited. *Prot. Sci.*, 17:1295–1307, 2008.

[53] Levitt, M.H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. John Wiley and Sons, 2001.

[54] Mittermaier, A. and Kay, L.E. Review - New tools provide new insights in NMR studies of protein dynamics. *Science*, 312(5771):224–228, 2006.

[55] Peng, J.W. and Wagner, G. Mapping of spectral density-functions using heteronuclear NMR relaxation measurements. *J. Magn. Reson.*, 98:308–332, 1992.

[56] Wang, T.Z., Frederick, K.K., Igumenova, T.I., Wand, A.J. and Zuiderweg, E.R.P. Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements. *J. Am. Chem. Soc.*, 127(3):828–829, 2005.

[57] Pelupessy, P., Ravindranathan, S. and Bodenhausen, G. Correlated motions of successive amide N-H bonds in proteins. *J. Biomol. NMR*, 25(4):265–280, 2003.

[58] Igumenova, T.I., Frederick, K.K. and Wand, J. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem. Rev.*, 106:1672–1699, 2006.

[59] Palmer, A.G., Kroenke, C.D. and Loria, J.P. Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. *Methods Enzymol.*, 339(Part B):204–238, 2001.

[60] Czerski, L., Vinogradova, O. and Sanders, C.R. NMR-Based amide hydrogen-deuterium exchange measurements for complex membrane proteins: Development and critical evaluation. *J.Magn.Reson.*, 142:111–119, 1999.

[61] Lewandowski, J.R., Sein, J., Blackledge, M. and Emsley, L. Anisotropic collective motion contributes to nuclear spin relaxation in crystalline proteins. *J. Am. Chem. Soc.*, 132(4):1246, 2010.

[62] Lewandowski, J.R., Sein, J., Sass, H.J., Grzesiek, S., Blackledge, M. and Emsley, L. Measurement of site-specific C-13 spin-lattice relaxation in a crystalline protein. *J. Am. Chem. Soc.*, 132(24):8252, 2010.

[63] Leavitt, S. and Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.*, 11:560–566, 2001.

[64] Brown, A. Analysis of cooperativity by isothermal titration calorimetry. *Int. J. Mol. Sci.*, 10:3457–3477, 2009.

[65] Grossoehme, N.E. and Giedroc, D.P. Energetics of allosteric negative coupling in the Zinc Sensor S. aureus CzrA. *J. Am. Chem. Soc.*, 131(49):17860–17870, 2009.

[66] Martinez-Julvez, M., Medina, M. and Velazquez-Campoy, A. Binding thermodynamics of ferredoxin:NADP(+) reductase: Two different protein substrates and one energetics. *Biophys. J.*, 96(12):4966–4975, 2009.

[67] Tochtrop, G.P., Richter, K., Tang, C.G., Toner, J.J., Covey, D.F. and Cistola, D.P. Energetics by NMR: Site-specific binding in a positively cooperative system. *Proc. Natl. Acad. Sci. USA*, 99(4):1847–1852, 2002.

[68] Protein Data Bank. http://www.rcsb.org.

[69] Branden, C. and Tooze, J. *Introduction To Protein Structure.* Garland publishing, 1998.

[70] Daune, M. *Molecular Biophysics, Structure in motion.* Oxford University Press, 1993.

[71] Wooll, J.O., Friesen, R.H.E., White, M.A., Watowich, S.J., Fox, R.O., Lee, J.C. and Czerwinski, E.W. Structural and functional linkages between subunit interfaces in mammalian pyruvate kinase. *J. Mol. Biol.*, 312(3):525–540, 2001.

[72] Bicout, D.J. and Zaccai, G. Protein flexibility from the dynamical transition: A force constant analysis. *Biophys. J.*, 80:1115–1123, 2001.

[73] Zaccai, G. How soft is a protein? A force constant approach to protein dynamics measured by neutron scattering. *Science*, 288:1604–1607, 2000.

[74] Balog, E., Becker, T., Oettl, M., Lechner, R., Daniel, R., Finney, J. and Smith, J.C. Direct determination of vibrational density of states change on ligand binding to a protein. *Phys. Rev. Lett.*, 93(2):028103, 2004.

[75] Consler, T.G., Uberbacher, E.C., Bunick, G.J., Liebman, M.N. and Lee, J.C. Domain interaction in rabbit muscle pyruvate-kinase 2. Small angle neutron scattering and computer simulation. *J.Biol.Chem.*, 263(6):2794–2801, 1988.

[76] Weiss, S. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature*, 7(9):724–729, 2000.

[77] Jares-Erijman, E.A. and Jovin, T.M. FRET imaging. *Nat. Biotech.*, 21(11):1387–1395, 2003.

[78] Klostermeier, D. and Millar, D.P. Time-resolved fluorescence resonance energy transfer: A versatile tool for the analysis of nucleic acids. *Biopolymers*, 61(3):159–179, 2001.

[79] Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.*, 86(4):2530 – 2537, 2004.

[80] Polit, A., Blaszczyk, U. and Wasylewski, Z. Steady-state and time-resolved fluorescence studies of conformational changes induced by cyclic AMP and DNA binding to cyclic AMP receptor protein from Escherichia coli. *Eur. J. Biochem*, 270:1413–1423, 2003.

[81] Hu, J., Ma, A. and Dinner, A.R. Monte Carlo simulations of biomolecules: the MC module in CHARMM. *J. Comput. Chem.*, 27:203–216, 2006.

[82] Du, Q., Smith, C., Shiffeldrim, N., Vologodskaia, M., Vologodskii, A. and Cozzarelli, N.R. Cyclization of short DNA fragments and bending fluctuations of the double helix. *Proc. Natl. Acad. Sci. USA*, 102(15):5397–5402, 2005.

[83] Thachuk, C., Shmygelska, A. and Hoos, H.H. A replica exchange Monte Carlo algorithm for protein folding in the HP model. *BMC Bioinf.*, 8:342, 2007.

[84] Oakley, M.T., Barthel, D., Bykov, Y., Garibaldi, J.M., Burke, E.K., Krasnogor, N. and Hirst, J.D. Search strategies in structural bioinformatics. *Curr. Protein Pept. Sci.*, 9:260–274, 2008.

[85] Case, D.A., Darden, T.A., Cheatham III, T.E., Simmerling, C., Wang, J., Duke, R., Luo, R., Crowley, M., Walker, R., Zhang, W., Merz, K., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, K., Paesani, F., Vanicek, J., Wu, X., Brozell, S., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D., Seetin, M., Sagui, C., Babin, V. and Kollman, P. Amber version 10. Technical report, University of California, San Francisco, 2008.

[86] Brooks, B.R., Brooks III, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M. and Karplus, M. CHARMM: The biomolecular simulation program. *J. Comp. Chem.*, 30:1545–1615, 2009.

[87] Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[88] Leach, A.R. *Molecular Modelling, principles and applications*. Prentice Hall, 2001.

[89] Louis, A.A. Beware of density dependent pair potentials. *J.Phys.Condens.Matter*, 14:9187–9206, 2002.

[90] Meirovitch, H., Cheluvaraja, S. and White, R.P. Methods for calculating the entropy and free energy and their application to problems involving protein flexibility and ligand binding. *Curr. Protein. Pept. Sci.*, 10(3):229–243, 2008.

[91] Miyamoto, S. and Kollman, P.A. Absolute and relative binding free-energy calculations of the interaction of biotin and its analogs with streptavidin using molecular-dynamics free-energy perturbation approaches. *Proteins Struct. Funct. Bioinf.*, 16(3):226–245, 1993.

[92] Patargias, G.N., Harris, S.A. and Harding, J.H. A demonstration of the inhomogeneity of the local dielectric response of proteins by molecular dynamics simulations. *J. Chem. Phys.*, 132:235103, 2010.

[93] Karplus, M. and Kushick, J.N. Method for estimating the configurational entropy of macromolecules. *Macromolecules*, 14(2):325–332, 1981.

[94] Andricioaei, I. and Karplus, M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, 115(14):6289–6292, 2001.

[95] Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.*, 215:617–621, 1993.

[96] Baron, R., van Gunsteren, W.F. and Hunenberger, P.H. Estimating the configurational entropy from molecular dynamics simulations: anharmonicity and correlation corrections to the quasi-harmonic approximation. *Trends Phys. Chem.*, 11:87–122, 2006.

[97] Harris, S.A. and Laughton, C.A. A simple physical description of DNA dynamics: quasi-harmonic analysis as a route to the configurational entropy. *J. Phys.: Condens. Matter*, 19(7), 2007.

[98] Bustamante, C., Bryant, Z. and Smith, S.B. Ten years of tension: single-molecule DNA mechanics. *Nature*, 421:423–427, 2003.

[99] Balsera, M.A., Wriggers, W., Oono, Y. and Schulten, K. Principal component analysis and long time protein dynamics. *J. Phys. Chem.*, 100(7):2567–2572, 1996.

[100] Grossfield, A., Feller, S.E. and Pitman, M.C. Convergence of molecular dynamics simulations of membrane proteins. *Proteins Struct. Funct. Bioinf.*, 67:31–40, 2007.

[101] Harris, S.A., Gavathiotis, E., Searle, M.S., Orozco, M. and Laughton, C.A. Cooperativity in drug-DNA recognition: A molecular dynamics study. *J. Am. Chem. Soc.*, 123:12658–12663, 2001.

[102] Stacklies, W., Xia, F. and Grater, F. Dynamic allostery in the methionine repressor revealed by force distribution analysis. *PLOS Comp.Biol.*, 5(11):1000574, 2009.

[103] Tirion, M.M. and Ben-Avraham, D. Normal mode analysis of G-actin. *J. Mol. Biol.*, 230:186–195, 1993.

[104] Tirion, M.M. Large amplitude elastic motions in proteins from single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77(9):1905–1908, 1996.

[105] Hinsen, K. Analysis of domain motions by approximate normal mode calculations. *Proteins Struct. Funct. Genet.*, 33(3):417–429, 1998.

[106] Lindahl, E., Azuara, C., Koehl, P. and Delarue, M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom Normal Mode Analysis. *Nucleic Acids Res.*, 34:W52–56, 2006.

[107] Liu, X. and Karimi, H.A. High-throughput modeling and analysis of protein structural dynamics. *Brief. Bioinform.*, 8(6):432–445, 2007.

[108] Bahar, I., Atilgan, A.R. and Erman, B. Direct evaluation o thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2:173–181, 1997.

[109] Flory, P.J. Statistical thermodynamics of random networks. *Proc. R. Soc. A*, 351:351–380, 1976.

[110] Delarue, M. Dealing with structural variability in molecular replacement and crystallographic refinement through normal-mode analysis. *Acta Crystallogr. D*, 64:40–48, 2008.

[111] Kundu, S., Melton, J.S., Sorensen, D.C. and Phillips, G.N. Dynamics of proteins in crystals: Comparison of experiment with simple models. *Biophys. J.*, 83(2):723–732, 2002.

[112] Yang, L.W., Rader, A.J., Liu, X., Jursa, C.J., Chen, S.C., Karimi, H.A. and Bahar, I. oGNM: online computation of structural dynamics using the Gaussian Network Model. *Nucleic Acids Res.*, 34:W24–W31, 2006.

[113] Bahar, I. and Rader, A.J. Coarse-grained normal mode analysis in structural biology. *Curr. Opin. Struct. Biol.*, 15:586–592, 2005.

[114] Tama, F., Gadea, F.X., Marques, S. and Sanejouand, Y.H. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. Genet.*, 41:1–7, 2000.

[115] Suhre, K. and Sanejouand, Y.H. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.*, 32(Suppl. 2):W610–W614, 2004.

[116] Williams, G. Elastic network model of allosteric regulation in protein kinase PDK1. *BMC Struct. Biol.*, 10(1):11, 2010.

[117] Ming, D. and Wall, M.E. Quantifying allosteric effects in proteins. *Proteins: Struct. Funct. Bioinf.*, 59:697–707, 2005.

[118] Jacobs, D.J., Rader, A.J., Kuhn, L.A. and Thorpe, M.F. Protein flexibility predictions using graph theory. *Proteins: Struct. Funct. Genet.*, 44:150–165, 2001.

[119] Wells, S., Menor, S., Hespenheide, B. and Thorpe, M.F. Constrained geometric simulation of diffusive motion in proteins. *Phys. Biol.*, 2:127–136, 2005.

[120] Micheletti, C., Lattanzi, G. and Maritan, A. Elastic properties of proteins: Insight on the folding process and evolutionary selection of native structures. *J. Mol. Biol.*, 321:909–921, 2002.

[121] Cooper, A., McAlpine, A. and Stockley, P.G. Calorimetric studies of the energetics of protein-DNA interactions in the Escherichia-coli methionine repressor (MetJ) system. *FEBS Lett.*, 348:41–45, 1994.

[122] Rafferty, J.B., Somers, W.S., Saint-Girons, I. and Phillips, S.E.V. Three-dimensional crystal structures of Escherichia coli Met repressor with and without corepressor. *Nature*, 341:705–710, 1989.

[123] Hawkins, R.J. *Coarse-grained models of dynamic allostery in proteins*. Ph.D. thesis, University of Leeds, 2005.

[124] Simonson, T. and Brooks III, C.L. Charge screening and the dielectric constant of proteins: Insights from molecular dynamics. *J. Am. Chem. Soc.*, 118:8452–8458, 1996.

[125] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. *Molecular biology of the cell*. Garland science, 2002.

[126] Toncrova, H. and McLeish, T.C.B. Substrate-modulated thermal fluctuations affect long-range allosteric signaling in protein homodimers: Exemplified in CAP. *Biophys. J.*, 98(10):2317–2326, 2010.

[127] Chaiancone, E., Vecchini, P., Verzili, D., Ascoli, F. and Antonini, E. Dimeric and tetrameric hemoglobins from the mollusc Scapharca inaequivalis. *J. Mol. Biol.*, 152:577–592, 1981.

[128] Koshland, D.E. The structural basis of negative cooperativity: Receptors and enzymes. *Curr. Opin. Struct. Biol.*, 6(6):757–761, 1996.

[129] Stevens, S.Y., Sanker, S., Kent, C. and Zuiderweg, E.R.P. Delineation of the allosteric mechanism of a cytidylyltransferase exhibiting negative cooperativity. *Nat. Struct. Biol.*, 8(11):947–952, 2001.

[130] Eaton, A.K. and Stewart, R.C. The two active sites of thermotoga maritima CheA dimers bind ATP with dramatically different affinities. *Biochemistry*, 48:6412–6422, 2009.

[131] Tsai, C.J., del Sol, A. and Nussinov, R. Protein allostery, signal transmission and dynamics: a classification scheme of allosteric mechanisms. *Mol. BioSyst.*, 5:207–216, 2009.

[132] Zhao, D., Arrowsmith, C.H., Jia, X. and Jardetzky, O. Refined solution structures of the Escherichia coli trp holo- and aporepressor. *J. Mol. Biol.*, 229(3):735–746, 1993.

[133] Yin, P., Fan, H., Hao, Q., Yuan, X., Wu, D., Pang, Y., Yan, C., Li, W., Wang, J. and Yan, N. Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat. Struct. Mol. Biol.*, 16(12):1230–U42, 2009.

[134] Strater, N., Hakansson, K., Schnappauf, G., Braus, G. and Lipscomb, W.N. Crystal structure of the T state of allosteric yeast chorismate mutase and comparison with the R state. *Proc. Natl. Acad. Sci. USA*, 93:3330–3334, 1996.

[135] Kong, Y.F., Ma, J.P., Karplus, M. and Lipscomb, W.N. The allosteric mechanism of yeast chorismate mutase: A dynamic analysis. *J. Mol. Biol.*, 356(1):237–247, 2006.

[136] Dev, I.K., Dallas, W.S., Ferone, R., Hanlon, M., McKee, D.D. and Yates, B.B. Mode of binding of folate analogs to thymidylate synthase - evidence for 2 asymmetric but interactive substrate-binding sites. *J. Biol. Chem.*, 269(3):1873–1882, 1994.

[137] Heyduk, T. and Lee, J.C. Escherichia Coli cAMP receptor protein: Evidence for three protein conformational states with different promoter binding affinities. *Biochemistry*, 28(17):6914–6924, 1989.

[138] Passner, J.M., Schultz, S.C. and Steitz, T.A. Modeling the cAMP-induced allosteric transition using the crystal structure of CAP-cAMP at 2.1 Angstrom resolution. *J. Mol. Biol.*, 304(5):847–859, 2000.

[139] Heyduk, E., Heyduk, T. and Lee, J.C. Intersubunit communications in Escherichia coli cyclic AMP receptor protein: studies of the ligand binding domain. *Biochemistry*, 31:3682–3688, 1992.

[140] Jarymowycs, V.A. and Stone, M.J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.*, 106:1624–1671, 2006.

[141] Li, L., Uversky, V.N., Dunker, A.K. and Meroueh, S.O. A computational investigation of allostery in the catabolite activator protein. *J. Am. Chem. Soc.*, 129:15668–15676, 2007.

[142] Popovych, N., Tzeng, S.R., Tonelli, M., Ebright, R.H. and Kalodimos, C.G. Structural basis for cAMP-mediated allosteric control of the catabolite activator protein. *Proc. Natl. Acad. Sci. USA*, 106:6927–6932, 2009.

[143] Leslie, A.G.W. and Wonacott, A.J. Structural evidence for ligand-induced sequential conformational changes in glyceraldehyde-3-phosphate dehydrogenase. *J. Mol. Biol.*, 178(3):743–772, 1984.

[144] Heddle, J.G., Okajima, T., Scott, D.J., Akashi, S., Park, S.Y. and Tame, J.R.H. Dynamic allostery in the ring protein TRAP. *J. Mol. Biol.*, 371(1):154–167, 2007.

[145] Hellmann, N., Weber, R.E. and Decker, H. Hemoglobin from the leech Macrobdella decora. *J. Biol. Chem.*, 278(45):44355–44360, 2003.

[146] Yonetani, T. and Laberge, M. Protein dynamics explain the allosteric behaviors of hemoglobin. *Biochim. Biophys. Acta*, 1784:1146–1158, 2008.

[147] Yonetani, T., Park, S., Tsuneshige, A., Imai, K. and Kanaori, K. Global allostery model of hemoglobin - Modulation of O-2 affinity, cooperativity, and Bohr effect by heterotropic allosteric effectors. *J. Biol. Chem.*, 277(37):34508–34520, 2002.

[148] Laberge, M. and Yonetani, T. Molecular dynamics simulations of hemoglobin a in different states and bound to DPG: Effector-linked perturbation of tertiary conformations and HbA concerted dynamics. *Biophys. J.*, 94(7):2737–2751, 2008.

[149] Richard, V., Dodson, G.G. and Mauguen, Y. Human deoxyhemoglobin-2,3diphosphoglycerate complex low-salt structure at 2.5 Angstrom resolution. *J. Mol. Biol.*, 233(2):270–274, 1993.

[150] Ackers, G.K. Deciphering the molecular code of hemoglobin allostery. In *Advances in protein chemistry*, volume 51, pages 185–253. Academic Press Inc., 1998.

[151] Brzozowski, A., Derewenda, Z., Dodson, E., Dodson, G., Grabowski, M., Liddington, R., Skarzynski, T. and Vallely, D. Bonding of molecular-oxygen to T-state human hemoglobin. *Nature*, 307(5946):74–76, 1984.

[152] Luisi, B., Liddington, B., Fermi, G. and Shibayama, N. Structure of deoxyquaternary hemoglobin with liganded with liganded beta-subunits. *J. Mol. Biol.*, 214(1):7–14, 1990.

[153] Jayaraman, V., Rodgers, K.R., Mukerji, I. and Spiro, T.G. Hemoglobin allostery - resonance raman spectroscopy of kinetica intermediates. *Science*, 269(5232):1843–1848, 1995.

[154] Yakushevich, L.V. *Nonlinear physics of DNA*. Wiley-VCH, 2004.

[155] Gavathiotis, E., Sharman, G.J. and Searle, M.S. Sequence-dependent variation in DNA minor groove width dictates orientational preference of Hoechst 33258 in A-tract recognition: solution NMR structure of the 2:1 complex with d(CTTTTGCAAAAG)(2). *Nucleic Acids Res.*, 28:728–735, 2000.

[156] von Hippel, P. From "simple" DNA-protein interactions to the macromolecular machines of gene expression. *Annu. Rev. Biophys. Biomol. Struct.*, 36:79–105, 2007.

[157] Hogan, M., Dattagupta, N. and Crothers, D.M. Transmission of allosteric effects in DNA. *Nature*, 278:521–524, 1979.

[158] Gilbert, D.E. and Feigon, J. Proton NMR-study of the [D(ACGTATACGT)]2-2 Echinomycin complex - conformational-changes between Echinomycin binding-sites. *Nucleic Acids Res.*, 20:2411–2420, 1992.

[159] Chenoweth, D.M. and Dervan, P.B. Allosteric modulation of DNA by small molecules. *Proc. Natl. Acad. Sci. USA*, 106(32):13175–13179, 2009.

[160] Panne, D., Maniatis, T. and Harrison, S.C. An atomic model of the interferon-beta enhanceosome. *Cell*, 129(6):1111–1123, 2007.

[161] Moretti, R., Donato, L.J., Brezinski, M.L., Stafford, R.L., Hoff, H., Thorson, J.S., Dervan, P.B. and Ansari, A.Z. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein - DNA assembly. *ACS Chem. Biol.*, 3(4):220–229, 2008.

[162] Pohl, F.M., Jovin, T.M., Baehr, W. and Holbrook, J.J. Ethidium bromide as a cooperative effector of a DNA structure. *Proc. Natl. Acad. Sci. USA*, 69:3805–3809, 1972.

[163] Koudelka, G.B. and Carlson, P. DNA twisting and the effects of noncontacted bases on affinity of 434 operator for 434 repressor. *Nature*, 355(6355):89–91, 1992.

[164] Rudnick, J. and Bruinsma, R. DNA-protein cooperative binding through variable-range elastic coupling. *Biophys. J.*, 76(4):1725–1733, 1999.

[165] Diamant, H. and Andelman, D. Binding of molecules to DNA and other semiflexible polymers. *Phys. Rev.E*, 61(6, Part B):6740–6749, 2000.

[166] Rappaport, S.M. and Rabin, Y. Model of DNA bending by cooperative binding of proteins. *Phys. Rev. Lett.*, 101(3), 2008.

[167] Coleman, S. *Aspects of symmetry*. Cambridge University Press, 1985.

[168] Woynarowski, J.M. AT islands - Their nature and potential for anticancer strategies. *Curr. Cancer Drug Targets*, 4(2):219–234, 2004.

[169] Searle, M.S. and Embrey, K.J. Sequence-specific interaction of Hoechst-33258 with the minor groove of an adenine-tract DNA duplex studied in solution by H-1-NMR spectroscopy. *Nucleic Acids Res.*, 18(13):3753–3762, 1990.

[170] Hamdan, I.I., Skellern, G.G. and Waigh, W.D. Use of capillary electrophoresis in the study of ligand-DNA interactions. *Nucleic Acids Res.*, 26(12):3053–3058, 1998.

[171] Rahimian, M., Miao, Y. and Wilson, W.D. Influence of DNA structure on adjacent site cooperative binding. *J. Phys. Chem. B*, 2008.

[172] Barkley, M.D. and Zimm, B.H. Theory of twisting and bending of chain macromolecules: analysis of the fluorescence depolarization of DNA. *J. Chem. Phys.*, 70:2991–3007, 1979.

[173] Watson, J.D. and Crick, F.H.C. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, 1953.

[174] Calladine, C.R. and Drew, H.R. *Understanding DNA - The molecule and how it works.* Academic Press, 1997.

[175] Perez, A., Noy, A., Lankas, F., Luque, F.J. and Orozco, M. The relative flexibility of B-DNA and A-RNA duplexes: database analysis. *Nucleic Acids Res.*, 32(20):6144–6151, 2004.

[176] Perez, A., Lankas, F., Luque, F.J. and Orozco, M. Towards a molecular dynamics consensus view of B-DNA flexibility. *Nucleic Acids Res.*, 36(7):2379–2394, 2008.

[177] Smith, S.B., Finzi, L. and Bustamante, C. Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads. *Science*, 258(5085):1122–1126, 1992.

[178] Smith, S.B., Cui, Y.J. and Bustamante, C. Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded DNA molecules. *Science*, 271(5250):795–799, 1996.

[179] Wang, M., Yin, H., Landick, R., Gelles, J. and Block, S.M. Stretching DNA with optical tweezers. *Biophys. J.*, 72:1335–1346, 1997.

[180] Bustamante, C., Smith, S.B., Liphardt, J. and Smith, D. Single-molecule studies of DNA mechanics. *Curr. Opin. Struct. Biol.*, 10:279–285, 2000.

[181] Matsumoto, A. and Go, N. Dynamic properties of double-stranded DNA by normal mode analysis. *J. Chem. Phys.*, 110(22):11070–11075, 1999.

[182] Lankas, F., Sponer, J., Hobza, P. and Langowski, J. Sequence-dependent elastic properties of DNA. *J. Mol. Biol.*, 299:695–709, 2000.

[183] Bustamante, C., Marko, J.F., Siggia, E.D. and Smith, S. Elasticity of λ-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[184] Marko, J.F. and Siggia, E.D. Bending and twisting elasticity of DNA. *Macromolecules*, 27:981–988, 1994.

[185] Marko, J.F. and Siggia, E.D. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.

[186] Nelson, P. *Biological physics; Energy, Information, Life.* W.H. Freeman, 2003.

[187] Odijk, T. Stiff chains and filaments under tension. *Macromolecules*, 28(7016-7018), 1995.

[188] Strick, T.R., Allemand, J.F., Bensimon, D., Bensimon, A. and Croquette, V. The elasticity of a single supercoiled DNA molecule. *Science*, 271(5257):1835–1837, 1996.

[189] Kahn, J.D. and Crothers, D.M. Protein-induced bending and DNA cyclization. *Proc. Natl. Acad. Sci. USA*, 89(14):6343–6347, 1992.

[190] Bednar, J., Furrer, P., Katritch, V., Stasiak, A.Z., Dubochet, J. and Stasiak, A. Determination of DNA persistence length by cryoelectron microscopy - separation of the static and dynamic contributions to the apparent persistence length of DNA. *J. Mol. Biol.*, 254(4):579–591, 1995.

[191] Okonogi, T., Reese, A.W., Alley, S.C., Hopkins, P.B. and Robinson, B.H. Flexibility of duplex DNA on the submicrosecond timescale. *Biophys. J.*, 77(6):3256–3276, 1999.

[192] Zhurkin, V. B. Ulyanov, N.B., Gorin, A.A. and Jernigan, R.L. Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA*, 88(16):7046–7050, 1991.

[193] Vologodskii, A.V., Levene, S.D., Klenin, K.V., Frankkamenetskii, M. and Cozzarelli, N.R. Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.*, 227:1224–1243, 1992.

[194] Olson, W. Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.*, 6:242–256, 1996.

[195] Beveridge, D.L. and McConnell, K.J. Nucleic acids: theory and computer simulation, Y2K. *Curr. Opin. Struct. Biol.*, 10:182–196, 2000.

[196] Qu, X.G., Ren, J.S., Riccelli, P.V., Benight, A.S. and Chaires, J.B. Enthalpy/entropy compensation: Influence of DNA flanking sequence on the binding of 7-amino actinomycin D to its primary binding site in short DNA duplexes. *Biochemistry*, 42(41):11960–11967, 2003.

[197] Lankas, F., Sponer, J., Langowski, J. and Cheatham, T.E. DNA basepair step deformability inferred from molecular dynamics simulations. *Biophys. J.*, 85:2872–2883, 2003.

[198] Olson, W.K., Gorin, A.A., Lu, X.J., Hock, L.M. and Zhurkin, V.B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA*, 95(19):11163–11168, 1998.

[199] Landau, L.D. and Lifshitz, E.M. *Theory of elasticity*. Elsevier, 1986.

[200] Leger, J.F., Robert, J., Bourdieu, L., Chatenay, D. and Marko, J.F. RecA binding to a single double-stranded DNA molecule: A possible role of DNA conformational fluctuations. *Proc. Natl. Acad. Sci. USA*, 95(21):12295–12299, 1998.

[201] Ali, B.M.J., Amit, R., Braslavsky, I., Oppenheim, A.B., Gileadi, O. and Stavans, J. Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proc. Natl. Acad. Sci. USA*, 98(19):10658–10663, 2001.

[202] Amit, R., Oppenheim, A.B. and Stavans, J. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophys. J.*, 84(4):2467–2473, 2003.

[203] Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham III, T., Laughton, C. and Orozco, M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.*, 92:3817–3829, 2007.

[204] Voth, G., editor. *Coarse-graining of condensed phase and biomolecular systems*. CRC Press, 2009.

[205] Dickerson, R.E. Definitions and nomenclature of nucleic-acid structure components. *Nucleic Acids Res.*, 17(5):1797–1803, 1989.

[206] Olson, W.K., Bansal, M., Burley, S.K., Dickerson, R.E., Gerstein, M., Harvey, S.C., Heinemann, U., Lu, X.J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.S., Westhof, E., Wolberger, C. and Berman, H.M. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, 313(1):229–237, 2001.

[207] Lu, X.J. and Olson, W.K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, 31(17):5108–5121, 2003.

[208] Landau, L.D. and Lifshitz, E.M. *Statistical physics, part 1.* Butterworth-Heinemann, Oxford, 3rd edition, 1980.

[209] Ming, D. and Wall, M.E. Allostery in a coarse-grained model of protein dynamics. *Phys. Rev. Lett.*, 95(19):198103, 2005.

[210] Amouyal, M. Gene regulation at-a-distance in E-coli: new insights. *C.R.Biologies*, 328(1):1–9, 2005.

[211] Schurr, J.M., Delrow, J.J., Fujimoto, B.S. and Benight, A.S. The question of long-range allosteric transitions in DNA. *Biopolymers*, 44(3):283–308, 1997.

[212] Pray, T.R., Burz, D.S. and Ackers, G.K. Cooperative non-specific DNA binding by octamerizing lambda cI repressors: A site-specific thermodynamic analysis. *J. Mol. Biol.*, 282(5):947–958, 1998.

[213] Wang, Q. and Calvo, J.M. LRP, a global regulatory protein of Escherichia-coli, binds cooperatively to multiple sites and activates transcription of ILVIH. *J. Mol. Biol.*, 229(2):306–318, 1993.

[214] Sun, L.J., Peterson, B.R. and Verdine, G.L. Dual role of the nuclear factor of activated T cells insert region in DNA recognition and cooperative contacts to activator protein 1. *Proc. Natl. Acad. Sci. USA*, 94(10):4919–4924, 1997.

[215] Julicher, F. Supercoiling transitions of closed DNA. *Phys. Rev. E*, 49(3):2429–2435, 1994.

[216] Kamien, R.D., Lubensky, T.C., Nelson, P. and O'Hern, C.S. Direct determination of DNA twist-stretch coupling. *Europhys. Lett*, 38(3):237–242, 1997.

[217] Humphrey, W., Dalke, A. and Schulten, K. Vmd - visual molecular dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[218] Magnus, J.R. and Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley-VCH, 1999.

[219] Wattis, J.A.D., Harris, S.A., Grindon, C.R. and Laughton, C.A. Dynamic model of base pair breathing in a DNA chain with a defect. *Phys. Rev. E*, 63(6, Part 1):art. no.–061903, 2001.

[220] Privalov, P.L., Dragan, A.I., Crane-Robinson, C., Breslauer, K.J., Remeta, D.P. and Minetti, C.A.S.A. What drives proteins into the major or minor grooves of DNA? *J. Mol. Biol.*, 365(1):1–9, 2007.

[221] Kirschner, K., Gallego, E., Schuster, I. and Goodall, D. Co-operative binding of nicotinamide-adenine dinucleotide to yeast glyceraldehyde-3-phosphate dehydrogenase.1. Equilibrium and temperature-jump studies at pH 8.5 and 40 degrees C. *J. Mol. Biol.*, 58(1):29–&, 1971.

[222] Shen, Y.Q., Li, J., Song, S.Y. and Lin, Z.J. Structure of apo-glyceraldehyde-3-phosphate dehydrogenase from Palinurus versicolor. *J. Struct. Biol.*, 130(1):1–9, 2000.

[223] Skarzynski, T. and Moody, P. C. E. and Wondacott, A. J. Structure of holo-glyceraldehyde-3-phosphate dehydrogenase from Bacillus-Stearothermophilus at 1.8 A resolution. *J. Mol. Biol.*, 193(1):171–187, 1987.

[224] Kitov, P.I., Shimizu, H., Homans, S.W. and Bundle, D.R. Optimization of tether length in nonglycosidically linked bivalent ligands that target sites 2 and 1 of a Shiga-like toxin. *J. Am. Chem. Soc*, 125(11):3284–3294, 2003.

[225] Yung, A., Turnbull, W.B., Kalverda, A.P., Thompson, G.S., Homans, S.W., Kitov, P. and Bundle, D.R. Large-scale millisecond intersubunit dynamics in the B subunit homopentamer of the toxin derived from Escherichia coli O157. *J. Am. Chem. Soc*, 125:13058–13062, 2003.