Unraveling the Complexity of Biological Processes from Protein Native Dynamics to Cell Motility in Molecular Simulations

Pavel I. Zhuravlev

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry.

Chapel Hill

2010

Approved by:

Advisor: Professor Garegin A. Papoian

Reader: Professor Michael Rubinstein

Reader: Professor Max Berkowitz

Reader: Professor Nikolay V. Dokholyan

Reader: Professor John Papanikolas

Abstract

Pavel I. Zhuravlev

Unraveling the Complexity of Biological Processes from Protein Native Dynamics to Cell

Motility in Molecular Simulations.

(Under the direction of Dr. Garegin A. Papoian)

This dissertation consists of two major parts. Both are dedicated to studying biological molecular processes with computer simulations, but differ in the scale of the studied processes. In one project we investigated the dynamics of cellular organelles involved in cell motility – the filopodia. The other project is zooming in to the scale of single molecule, elucidating the organization of protein molecule native state.

Some motile cells use special fingerlike probes of their environment for guiding their motion called *filopodia*. They are bundles of parallel actin filaments protruding from the cell body and enveloped by cell's membrane. They are highly dynamic, constantly growing an retracting, randomly, or in response to the change in the environment. These dynamics are governed by the cell's regulatory proteins and by external chemical cues or mechanical obstacles. The previous models predicted that a filopodium grows to a stationary length of about 1 μm with miniscule fluctuations around. (i) We found that capping proteins (they attach to the barbed ends of actin filaments and stop polymerization) can induce macroscopic oscillation of filopodial length – the growth-retraction cycles. The retraction can be complete. This is the first model that predicts finite lifetimes for filopodia. The lifetimes are consistent with experimental observations. (ii) In the model, however, the maximal filopodial lengths of several microns

are still limited by the diffusional transport of actin monomers to the filopodial tip and are far below experimentally observed lengths of up to 100 μm. Assuming the obvious solution for the problem of slow transport in cell, the molecular motors, that are known to be present inside filopodia, we found that a naïve addition of motors does not increase the lengths much. In order to have an efficient active transport, two rules must be observed: the motors should not sequester the cargo and the rails for motors should be kept from being clogged by motors. Protein Ena/VASP that is known to be actively transported to the filopodial tip by molecular motors may be a way to fight sequestration.

On the scale of a single macromolecule we studied the organization of protein native state. It is not a single structure, but an ensemble of constantly interconverting conformations. It is essential for a deep insight into protein functioning to know thermodynamics of these substates and dynamical regime of their exploration. (i) In all-atom MD simulations we constructed a 2D free energy surface for a protein Trp-cage and using the FES for Brownian dynamics investigated the nature of dynamical behavior of Trp-cage in its native state. We found that the dynamical regime is borderline between liquid and supercooled liquid. (ii) We developed a general technique for calculating free energy difference between two polymer conformations in explicit solvent simulations and used the Trp-cage 2D FES for testing of this technique, revealing remarkable accuracy and computation efficiency.

Acknowledgements

It is a pleasure to thank those who made this dissertation possible. I am most deeply grateful to my advisor, Dr. Garyk Papoian, who, first of all, encouraged me to enroll to this graduate program and then provided a persistent intellectually enlightening guidance during all the years. He successfully created environments for fruitful brainstorming sessions and discussions, allowing sufficient freedom of research on my own, but at the same time turning my attention off obviously stray paths.

This dissertation would not be possible without the members of our research group, who were providing a lot of help and incessant ideas in countless enlightening discussions, especially, Dr. Yueheng Lan, Dr. Alexey Savelyev and Chris Materese, who also spent a lot of his time editing the texts I produced.

I am grateful to people who organized and participated in Polymer Physics Seminar, which were very intellectually stimulating. Various journal and reading discussion clubs organized by Dr. Michael Rubinstein, Dr. Max Berkowitz, Dr. Nikolay Dokholyan and their research groups have considerably contributed to my shaping as a scientist.

I thank people in Chemistry Department and various departments of UNC Medical School who organized wonderful seminars inviting brilliant and prominent speakers. This scientific environment was invaluable for my research.

I am thankful to my parents who from the very beginning fully supported the idea of going to graduate school in another country.

There are also people who greatly contributed specifically to the research presented in this dissertation. I am grateful to Dr. Richard Cheney and his research group members for helpful

discussions about possible roles of Myosin X and Ena/VASP in filopodia. I thank Dr. Michael Rubinstein for many stimulating discussions on the protein dynamics project.

Table of Contents

x

List of Figures

List of Tables

List of Abbreviations and Symbols

1D, 2D, 3D   One-, two-, three-dimensional

$Å$ Ångström, 10^{-10} m

ATP, ADP   Adenosine triphosphate, adenosine diphosphate

DSM   Dielectric Solvent Model

Ena/VASP   Enabled/vasodilator-stimulated phosphoprotein

F-actin   Filamentous actin, actin polymer

FES   Free energy surface

FPT   First-passage time

G-actin   Globular actin, actin monomers

k, k_B Boltzmann constant

K   Kelvin

kcal/mol   Kilocalories per mole

LAMMPS   Large-scale Atomic/Molecular Massively Parallel Simulator

μM   micromolar, 10^{-6} mol/L

μm   micrometer, 10^{-6} m

μs   microsecond, 10^{-6} s

M10	Myosin X molecular motor
MD	Molecular dynamics
MFPT	Mean first-passage time
ms	millisecond, 10^{-3} s
nM	nanomolar, 10^{-9} mol/L
nm	nanometer, 10^{-9} m
NMR	Nuclear Magnetic Resonance
NPT	Isothermal-Isobaric ensemble
ns	nanosecond, 10^{-9} s
PDB	Protein Data Bank
RMSD	Root-mean-square deviation
T	Temperature
v_{retr}	Retrograde flow velocity
WHAM	Weighted Histogram Analysis Method

Chapter 1

Introduction

1.1   Cell motility and filopodia

Eukaryotic cells that have to move to perform their functions rely on actin based organelles for motility [1]. On a surface, these cells move by extending a flat layer with a three dimensional actin mesh inside called a lamellipodium. Thickness of a lamellipodium is on the order of 100 nm and transverse dimensions are above the micron length scale. The front of an extending lamellipodium is called the leading edge. From the leading edge the cell extends thinner finger-like protrusions called filopodia [2]. They emerge from the lamellipodium, at those points of the leading edge where actin filaments of the branched 3D mesh bundle together in parallel. Primary role of filopodia is to sense the environment and help to guide the cell's motion.

The first part of this dissertation focuses on modeling various aspects of filopodial growth dynamics. Filopodia are employed by fibroblasts – motile cells that perform wound-healing – while these cells are finding its way through the body to the wound in order to cover it [3]. During embryo formation, neurons grow axons, which may be up to a meter long. A front of a growing axon – a neural growth cone – pauses every now and then extending multiple filopodia to probe the surroundings and decide in which direction it has to turn [4]. Filopodia

are needed for dorsal closure in drosofila embryos [5] and are implicated in cancer metastatic propagation [6].

Extension and retraction of filopodia and lamellipodia are based on actin polymerization and depolymerization which are, in their turn, affected by various regulatory proteins [7]. Actin filament (F-actin) is asymmetric, and the polymerization-depolymerization rates are different on the two ends called *the barbed end* and *the pointed end*. In a living cell, typically, the polymerization at the barbed end is considerably faster, so a filament can be thought of as growing from the barbed end. Barbed ends in lamellipodia and filopodia are near the cell's membrane pushing it forward during polymerization. In lamellipodia, the branching agent, Arp2/3 can attach and start a new filament at an angle to the existing one [8], thus forming the 3D actin mesh. Another important regulator is capping protein that attaches to the barbed end and stops polymerization completely until it falls off. Concentration of actin monomers (G-actin) near barbed ends obviously influences polymerization and growth rates. Amount of free G-actin is regulated by special sequestering proteins [9]. A filopodium emerges when lamellipodial actin filaments group together in parallel at a specific spot of the leading edge, increasing there the pressure on the membrane and starting the protrusion. The pointed ends of filopodial filaments remain in the lamellipodial actin mesh. Apart from capping and sequestering proteins, there are many others that regulate the filopodial growth. Anticapping proteins prevent capping proteins from attaching to the barbed ends, they also promote polymerization [10]. Fascins and other cross-linking proteins bundle the parallel filaments together increasing mechanical stability of the filopodium [11]. These and other multiple regulatory proteins form an intricate chemical signaling network.

As a probe, a filopodium is sensitive to changes in environment and to fluctuations in the regulatory network. It switches from growth to retraction and back to growth in response to these internal fluctuations or external influences, such as chemical cues or mechanical obstacles. The process is known as growth – retraction cycles. The aforementioned chemical

network is tightly coupled to mechanics. Elastic force from the membrane pushes back the polymerizing barbed ends at the filopodial tip. Apart from that, active motions, which are based on myosin molecular motors, in the lamellipodial actin mesh pull the filaments into the cell [12]. The total effect of these two (and possibly more) processes is known as retrograde flow. Competing with polymerization, it engenders the complicated growth-retraction dynamical behavior. Actin filaments are semiflexible polymers with persistence length on the order of $20\,\mu$m [13]. Under the load they will buckle, though, due to cross-linking the buckling force for the filopodial bundle of filaments will be larger. In addition, in the cells moving over a surface lamelliodial and filopodial F-actin can attach to the substrate in the specific spots called focal adhesions [14].

Summarizing, filopodial dynamics is a complicated mechano-chemical system representing a challenge for both theoretical and experimental studying. Modeling of this mechano-chemical signaling network assists in capturing the essential effects of regulating various proteins on the network dynamics, where corresponding experiments would require an immense effort.

1.1.1 Stochastic simulations of mechano-chemical networks

Chemical part of the signaling networks that regulate filopodia consists of proteins and their interactions through chemical reactions. In many cases a set of chemical reactions is analyzed through writing down a system of ordinary first-order differential equations with time as independent variable and concentrations of interacting species as dependent variables, that is, functions of time. The description of the chemical network based on the solution to this system of equations is known as chemical kinetics description. The concentrations in these equations are average numbers of molecules in a unitary volume. However, chemical reactions are discrete random processes. Reactants float in the solution or in the gas phase and have to encounter each other randomly, but even unary reactions, such as radioactive decay are random

events. When the numbers of reacting molecules are large, on the order of Avogadro's number, the relative fluctuations of these numbers are negligible. In such cases, time evolution of averages gives an appropriate description of the system dynamics, and chemical kinetics can be used.

A crucial feature of biological signaling networks is that the average numbers of molecules of each reacting protein in the relevant spatial region are very low, on the order of several molecules (down to 10^{-2} molecules for some of the processes studied in this dissertation). In this case, the fluctuations (estimated as a square root of number of molecules) are on the same scale as the average and can even exceed it by an order of magnitude. Chemical kinetics may not provide a physically meaningful answer for a biological signaling network, therefore, the dynamics of such network has to be treated stochastically.

In order to illustrate the difference between chemical kinetics and stochastic approaches let us consider a unary reaction where species A can convert to B with rate k_1 and backwards with rate k_2:

$$A \underset{}{\overset{k_1,\ k_2}{\rightleftharpoons}} B. \tag{1.1}$$

Instead of following evolution of concentrations $[A]$ and $[B]$, the stochastic approach treats the dynamics of this chemical system as a random walk on a 1D lattice with nodes corresponding to particular copy numbers for each species involved in the chemical network. If the system starts with n molecules of A and m molecules of B (state (n,m)), one of the A molecules can convert to B with rate k_1 per molecule, and one of the B molecules can convert to A with rate k_2 per molecule. This translates to a step to the right along the lattice with rate nk_1 or step to the left with rate mk_2.:

$$\ldots \underset{}{\overset{(n+2)k_1,\ (m-1)k_2}{\rightleftharpoons}} (n+1,m-1) \underset{}{\overset{(n+1)k_1,\ mk_2}{\rightleftharpoons}} (n,m) \underset{}{\overset{nk_1,\ (m+1)k_2}{\rightleftharpoons}} (n-1,m+1) \underset{}{\overset{(n-1)k_1,\ (m+2)k_2}{\rightleftharpoons}} \ldots \tag{1.2}$$

The complete description of this random walk is given by the probability distribution

$P(n,m,t)$ – the probability that the system is in the node (n,m) at time t. This function is a solution of the so-called Master Equation.

The example is zero-dimensional, in the spatially resolved case, the average concentrations also depend on spatial coordinates, and chemical kinetics equations are complemented by diffusion equations. Stochastic spatially resolved system has a lattice similar to (1.2) at each point in space. In case of signaling network regulating filopodial growth, the lattice is highly multidimensional, spatially resolved, and in addition, the rates for steps between the nodes are dependent on mechanical degrees of freedom, such as retrograde flow velocity, membrane position or bending and buckling of the filaments. Solving of this equation analytically is hopeless, but stochastic computer simulations may be employed.

In the model used in this dissertation, we discretize space into compartments and keep track of copy numbers of all chemical species in all compartments. To characterize the distribution function, which is the solution of the master equation, we run multiple realizations of the random process and calculate averages, variances, joint distributions etc. This approach is similar to running multiple Langevin trajectories to obtain characteristics of probability distribution, which is a solution of corresponding Fokker-Planck equation. The technique may be called, therefore, chemical Langevin dynamics simulations.

The implementation of such Langevin dynamics of basic filopodial mechano-chemical network has been developed in our research group based on the Gillespie algorithm for stochastic simulations [15]. Gillespie algorithm draws random numbers to decide which reaction and where in space occurs next, thus realizing the stochastic propagator on the lattice of chemical network. The basic model for the filopodial growth only includes diffusion of actin monomers, polymerization, retrograde flow and mechanical influence of the membrane. No regulatory proteins are considered in the basic model. It predicts growth to a stationary length determined by balance of diffusional flux of actin monomers forward and retrograde flow actin flux backward. In this dissertation we build on the basic model [15], adding regulatory proteins, to investigate

their influence on the filopodial lengths and growth speeds.

1.1.2 Capping protein noise amplification

Chapter 2 discusses capping proteins, which are present in filopodia in very low concentrations. If a filament is capped, it starts to retract due to retrograde flow. When all filaments are capped, the whole filopodium switches to retraction. Capping proteins fall off from the barbed ends at a slow rate, so when enough filaments are uncapped to overcome the membrane force, the filopodium switches back to growth. Eventually, at some retraction phase, the filopodium would retract all the way to the cell body and disappear, showing now a finite lifetime instead of growing to stationary length. The filopodial lifetimes of several minutes on average, obtained in our simulations agree with experimentally known values. Low concentration of capping proteins means that capping process is a discrete slow random noise, and a random switch between two fast processes of growth and retraction, and the filopodium thus acts as an amplifier of that noise, making it visible on macroscopic temporal (~ 100 s) and spatial (~ 100 nm) scales. This high susceptibility to tiny fluctuation may be beneficial to sensorial role of filopodia. The second part of Chapter 2 presents analytical estimation for individual filament disappearance rates based on mean capping and uncapping times.

1.1.3 Active transport in filopodia

The bottleneck for filopodial length is the diffusional flux of G-actin to the filopodial tip, which decreases with the length [15]. A "standard" biological solution such a problem, lack of performance in diffusional transportation, is the use of molecular motors. The picture of such active transport suggested by cartoons in biology textbooks, shows cargo loaded onto motors which walk forward and unload cargo at the destination, much like a conveyor belt. In fact, Myosin X molecular motors, which can walk along actin filaments, have been observed inside filopodia and shown to influence filopodial formation [16]. Chapter 3 investigates the possibility of

6

G-actin being transported by these motors to the filopodial tip. It turns out that a mere addition of these motors to the mechano-chemical network does not affect the transport efficiency a lot, because motors sequester the cargo. Intricate schemes, involving multifunctional proteins may work better. For instance, Ena/VASP is known to be transported by Myosin X inside filopodia [17] and also cross-links the actin filaments near the tip [11]. Assuming that Ena/VASP is needed as an adaptor between G-actin and Myosin X and consumed for cross-linking near the tip, we found that these two features can prevent sequestration of G-actin and make transport more efficient. Still, the conveyor-like transportation is only achieved in a narrow ranges of parameters.

1.2   Energy landscape perspective on the dynamics of the protein native state

Much more often, molecular simulations in biology are used on a considerably smaller scale. If we zoom into the scale of a single protein molecule, we find that it is dynamic and complicated. Protein interactions and rate constants for the reactions between proteins, which are the input parameters in the first part of the dissertation, follow in fact from protein structural and dynamical organization. In the second part of the dissertation we develop techniques permitting to study this organization.

A few decades ago, the question of how an amino acid sequence can predefine a unique 3D folded structure for a protein seemed hopelessly intractable. Even then, however, kinetic models were successfully devised to describe the conformational transitions in allosteric proteins such as hemoglobin. Subsequently, the statistical view of the protein folding problem lead to a realization that globular proteins are built in a very special way, which tremendously simplifies the folding problem. This, in turn, has allowed nature to evolve complex protein folds.

The energy landscapes of foldable proteins are funneled – the conformations that are struc-

turally similar to the native state are also low in energy, while the native state interactions are minimally frustrated [18, 19], diminishing the energetic ruggedness which could kinetically prevent folding [20–30]. At the same time, the native state, that used to be thought of as a single structure – simple and clear – turned out to be a collection of conformations that are explored as the natively folded protein roams the landscape at the bottom of the funnel [31–36]. This portion of the landscape (that we will afterwards call a functional landscape) has a much smaller phase volume compared with the whole protein folding phase space, but is much richer in its topography when viewed at high resolution – in other words it is more rugged and glassy. Ironically, the folding landscape that has astronomical number of states turned out to be easier to understand conceptually, than this small region at the bottom of the funnel. Despite a number of open questions still remaining in protein folding, energy landscape theory and the funnel paradigm ushered in a general understanding and provided a common language, which is used widely by experimentalists and theoreticians [37–41]. With the functional landscapes, however, it is likely that most of the properties of protein functional dynamics depend on the particular features of the landscape of a specific protein, which might prevent emergence of a general universal view. However, the energy landscape language can still provide a fruitful common ground for discussion of protein function and dynamics. These dynamics represent motion in multidimensional manifold riddled with cliffs, ridges, peaks and crevices (also multidimensional). For deep understanding of these phenomena, we need a reduced representation of the protein energy landscape and of it exploration by the protein molecule. Or, to put it in form of simple questions: How many variables are needed to describe the relevant motions? What are these variables and how to systematically find them?

Because of its minimally frustrated nature, the folding landscape produces kinetics that are often adequately mapped to a single coordinate, the fraction of native contacts, denoted as Q [42]. The average energy of contacts decreases as a function of Q, so the energy landscape may be visualized as a funnel (Fig. 1.1). These statements are not only qualitatively

appealing, but are also quantitative, based on the generalized random energy model (GREM), which was borrowed from spin glass physics [43]. In the middle part of the funnel, the protein chain is already compact, but very dynamic, still possessing significant entropy [44–47]. One consequence of this is the robust self-averaging property: an ensemble of energies sampled on a small part of the landscape is to a large extent representative of the whole ensemble of energies [48–50]. Self-averaging permits use of the random energy model (REM) which is, in essence, replacing the detailed description of a particular landscape (like a detailed topographical map) with an estimated probability distribution of energy minima and barriers. Configurational entropy in REM is determined by the interplay between total number of possible states and the landscape ruggedness. Therefore, if the funnel is stratified at various Q values, one may compute the configuration entropy as a function of Q, at the given temperature (see Fig. 1.1). Using reasonable parameters to describe the funnel, it was found that at $Q \gtrsim 0.7$ the configurational entropy is severely diminished, indicating a transition to more glassy dynamics [19]. Importantly, the convenient self-averaging character of low Q strata is no longer observed for many dynamical variables during the interconversion of nearly folded states. Therefore, the bottom of the funnel must be described with a detailed map of the functional energy landscape. Q by itself may no longer be the main order parameter, as it is correlated with the folding, that has already occurred, while we are interested in functional motions of the folded protein. Therefore, coordinates transverse to Q play an important role. On the positive side, due to small phase volume of the native state ensemble, a brute-force sampling of the functional landscape is feasible, as apparent from numerous experiments and actual protein functioning in biological organisms. One reason for that is that despite the functional landscape being glassy, the protein does not have to actually overcome barriers higher than the one associated with unfolding. From any point of the landscape it can unfold and refold to another point, so the distribution of the barriers (and therefore, of timescales) is capped. Indeed, typical excitations of a protein in native state include partial unfolding [51]. For the same reason, Q is still likely to be a relevant

coordinate, but more coordinates are required to describe the motions transverse to Q.

In the most general case, the functional energy landscape has multiple minima, which are hierarchically organized with a complicated transitions kinetics [21, 33]. The ultimate goal for computational modeling is to find an approach that is able to predict other native-like states and the kinetics of transitions between all of these states starting from an experimentally determined native structure.

As previously mentioned, the native protein dynamics are immensely multidimensional with several hundred protein residues corresponding to several thousand degrees of freedom, in addition to many more solvent degrees of freedom. However, solvent degrees of freedom are separated by a large timescale gap from the protein conformational motions. For this reason, when discussing the latter, it is often assumed that the solvent degrees of freedom are adiabatically equilibrated around any protein conformation and are present in the description in the implicit way. We follow this convention here: when discussing energy landscapes, by energy, we actually mean free energy of contacts, where solvent degrees of freedom are integrated out. On the other hand, conformational entropy remains explicit and is separate from this "effective energy". Even in this case, the energy landscape remains highly multidimensional and native dynamics cannot be interpreted unless projected on a low number of coordinates or represented in a different simplified way.

Proteins possess many structural components: some rigid, some flexible, and some liquid. These components move on a multitude of timescales ranging from picoseconds to seconds [52]. Likewise, the localization of these motions varies from atomic vibrations to global motions of the whole molecule [53]. When addressing a particular problem, motions on specific timescales and localizations are usually of interest. Functional allosteric motions, for instance, are typically in the slower range of the possible timescales and are spatially global. This places limitations on the set of motions investigated to describe a particular phenomenon and motivates the projection of the dynamics into a much lower dimensional space. Indeed,

techniques like principal component analysis (PCA) or Essential Dynamics are able to describe much of the non-trivial motion by a mere 3-5% of the total conformational degrees of freedom [54–58]. However, the essential phase subspace may be complicated, even fractal, so it is not clear how to choose these coordinates, starting from a protein crystal structure. Can there be a universal method for characterizing the essential subspace, or dissimilar approaches are better suited for different protein classes and sizes?

1.2.1 Free energy surfaces and differences as a way to represent the energy landscape

One approach is to directly derive coordinates from the spatial motions of the protein of interest, where these motions are either obtained experimentally, or from computer simulations [52] or are speculated based on the protein molecular architecture. Often these motions are connected to function, such as binding of oxygen to hemoglobin or ion channel opening and closing. If there is no clear way to correlate the coordinates with functional motions, they may be at least based on the structures of various functional states of the protein. The aforementioned quaternary functional motions, such as respective rearrangements of domains and subunits in the case of multidomain proteins and oligomers, are usually more obvious. However, smaller, single domain globular proteins may perform more subtle allosteric switching. In this case, good coordinates may be ones that correlate with minima or valleys in the energy landscape. At the very least, a good coordinate must evolve slowly on the timescale of interest.

In the second part of the dissertation we focus on another type of approach, the techniques that can create a representation of a protein's energy landscape which allows subsequent studying of specific questions, such as the nature of the dynamics (diffusive or activated). One approach is reconstructing the free energy surfaces of the native state. To enhance the coordinate Q we choose two different structures (A and B) from the native state and calculate free energy of the protein molecule as a function of two Qs now. For any given conformation, Q_A

is a fraction of contacts shared with structure A and Q_B is the fraction of contacts shared with B. Since both A and B are from the native ensemble and similar, both Q_A and Q_B are correlated with folding and can describe partial unfolding excitations (cracking). At the same time, together, they resolve the motions transverse to folding and allow multiple pathways unlike a 1D free energy profile. We tested this technique on a small protein Trp-cage. The free energy surface (FES) as a function of two Qs, obtain from all-atom explicit solvent molecular dynamics simulation, allowed us to investigate the dynamical regime of the Trp-cage native state at room temperature. Using the FES to run Brownian dynamics, we found that the dynamical behavior of Trp-cage is borderline between diffusive, or liquid-like, dynamics, and activated dynamics, characteristic of supercooled liquids. Another feature of a free energy surface technique is that it provides a way to compare different force-fields by comparing free energy surface produced by them. These types of comparisons can be used in developing coarse-grained force fields. We constructed the same FES for Trp-cage in "dielectric solvent" force-field, where water molecules were not present and electrostatic terms diminished 80-fold. The correspondence between the two free energy surfaces is surprisingly good, especially, in the most folded region. These results are reported in Chapter 4.

The construction of 2D free energy surface requires a lot of computational resources. As a first step towards building a representation of energy landscape it is possible to calculate just the free energy difference between two conformations. These also might be used in developing coarse-grained force fields and elucidating some biologically related conformational transitions. In Chapter 5 we report a fully general technique that can calculate free energy difference between two specific conformations of a polymer chain in explicit solvent computer simulations. We tested this method on the same molecule, Trp-cage, confirming the accuracy of the method within 5% through comparing to a more expensive 2D technique.

Summarizing, this dissertation reports the results of several computer simulation works of biological systems ranging from microscopic single molecule scale to mesoscopic scale of cel-

lular organelles. We have developed novel techniques for studying protein's native state, based on the energy lanscape paradigm, and created the most elaborate to date stochastic physico-chemical model of filopodial growth dynamics. We investigated role of capping proteins in filopodia, finding that they induce macroscopic instabilities in filopodial length by generating slow discrete random noise amplified by the filopodium. We pursued the possibility that actin monomers are actively transported inside filopodia, formulating two rules of efficient active transport, and coming up with a plausible biological scheme that can uphold one of these rules.
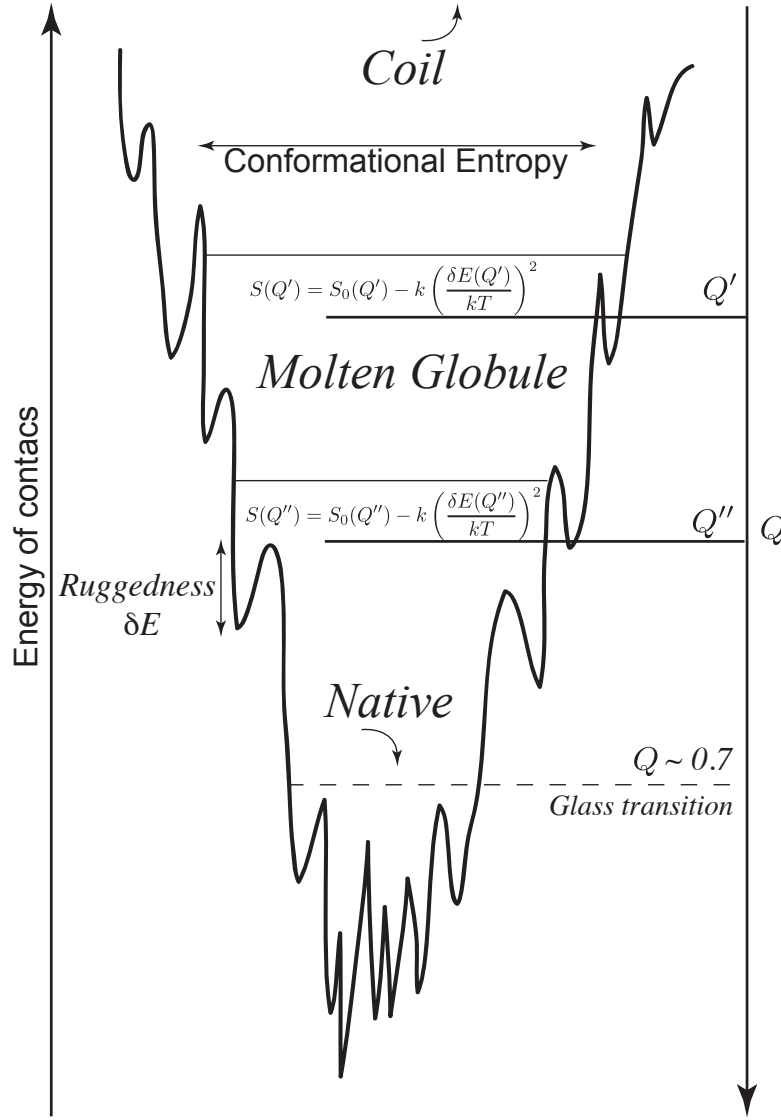
Figure 1.1: **The protein folding landscape.**

The folding landscape be schematically drawn as a funnel [19]. The vertical axis corresponds to energy of contacts, which is correlated to Q – the similarity to the native state. In the upper region the protein chain is unfolded with a large conformational entropy. In the middle part of the funnel, the molten globule region, the chain is compact, but retains significant entropy. The lower part is a collection of similar low-energy conformations separated by barriers, known as the native state. The funnel can be stratified according to Q. At a fixed Q, configurational entropy is defined by the interplay between total number of states and ruggedness of the landscape for that particular Q stratum. Glass transition temperature T_g corresponds to vanishing of the configurational entropy. In the figure, $T_g(Q'') > T_g(Q')$. The funnel region around $Q \sim 0.7$, which corresponds to the onset of native like conformations, is estimated to have a high glass transition temperature [19]

14

Chapter 2

Molecular Noise of Capping Protein Binding Induces Macroscopic Instability in Filopodial Dynamics

Pavel I. Zhuravlev and Garegin A. Papoian

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599

2.1 Abstract

Capping proteins are among the most important regulatory proteins involved in controlling complicated stochastic dynamics of filopodia, which are dynamic finger-like protrusions used by eukaryotic motile cells to probe their environment and help guide cell motility. They attach to the barbed end of a filament and prevent polymerization, leading to effective filament retraction due to retrograde flow. When we simulated filopodial growth in the presence of capping proteins, qualitatively new dynamics emerged. We discovered that molecular noise due to capping protein binding and unbinding leads to macroscopic filopodial length fluctuations, compared with minuscule fluctuations in the actin-only system. Thus, our work shows for the first time that molecular noise of signaling proteins may induce micron-scale growth-retraction cycles in filopodia. When capped, some filaments eventually retract all the way down to the filopodial base and disappear. This process endows filopodium with a finite lifetime. Additionally, the filopodia transiently grow several times longer than in actin-only system, since less actin transport is required due to bundle thinning. We have also developed an accurate mean field model which provides qualitative explanations of our numerical simulation results. Our results are broadly consistent with experiments, in terms of predicting filopodial growth retraction cycles and the average filopodial lifetimes.

2.2 Introduction

Eukaryotic motile cells project finger-like protrusions, called filopodia, to probe their environment and help guide cell motility [2]. They play important roles in neuronal growth [4], wound healing [3] and cancer metastasis [6]. The filopodial structure consists of parallel actin filaments, cross-linked into bundles by actin-binding proteins, all of which is enclosed by the cell's plasma membrane [2, 59]. Despite their importance in eukaryotic biology and human health, the physical mechanisms behind filopodial regulation and dynamics are poorly understood,

16

including what drives ubiquitous growth-retraction cycles and eventual filopodial disappearance [2–4, 6, 59–61]. Using stochastic simulations of filopodial dynamics, we have discovered that molecular noise due to binding/unbinding of a capping protein results in macroscopic growth-retraction fluctuations, compared with minuscule fluctuations in the actin only system. Due to rare fluctuations some filaments eventually retract all the way down to the filopodial base and disappear. In contrast to prior computational models that predicted stable filopodia at steady state, our simulations show that filopodial lifetimes are finite. We have also developed an accurate mean-field model which provides novel insights into filament disappearance kinetics.

A comprehensive computational model of a filopodium should contain the following features: mechanical interactions including membrane dynamics, protrusion force, and retrograde flow; chemical interactions, including actin polymerization and depolymerization; and biological signaling interactions that control the filopodia dynamics turnover. The first mean-field model for filopodial growth addressed the interplay between filament growth and diffusional actin transport [62]. A subsequent work highlighted the importance of the interactions between the membrane and filament barbed ends [63]. Our own previous study was the first to treat both polymerization and diffusion in a fully stochastic fashion [15]. In that model, the filopodia grow to some steady state length, and subsequently exhibit only slight fluctuations [15]. Thus, no essential dynamics occurs after the steady state is reached, a finding similar to those of other prior filopodial simulations [62, 63], implying an essentially infinite filopodial lifetime and no turnover.

Although it is not well known whether the turnover is driven externally or internally, it is plausible that internal biochemical reaction network dynamics is a significant contributor. For instance, capping proteins bind to the barbed ends of actin filaments preventing polymerization [64]. Their efforts are countered by formins [10], anticapping processive motors that attach to the barbed ends and may effectively increase polymerization rate up to five-fold [10].

In this work, we investigated the influence of two regulating proteins on the filopodial turnover process. The mechano-chemical model that we use here to describe the filopodial dynamics is fully stochastic, using the Gillespie algorithm to calculate simulation steps (see Fig. 2.1). It consists of the following processes: 1) the diffusion of proteins from the cytosol at the filopodial base to the tip; 2) the force applied by the membrane on individual filaments; 3) the actin filament polymerization and depolymerization at the barbed end; 4) the depolymerization at the pointed end and the induced retrograde flow v_{retr} of the filopodium as a whole; 5) capping of the filaments that stops polymerization [64]; 6) binding of formin to the barbed end that increases effective polymerization rate fivefold [10]. The filopodium is split into compartments with diffusion realized as stochastic hops between them. Membrane force is taken into account by effectively decreasing the polymerization rates (see Methods Secton and Ref. [15] for more details). The parameters, such as reaction rates and concentrations, are given in Table 2.1.

2.3   Results and Discussion

The addition of capping proteins and formins results in very different, more complex growth dynamics (see Fig. 2.2), in contrast to actin-only model in Ref. [15], where the observed length of the filopodium was stationary. In particular, filopodial length fluctuations become *macroscopic*, increasing from below 100 nm in a model without these proteins to a few microns, on the order of the length of filopodia. In addition, a clearly identifiable retraction phase appears, with a lifetime of ~100 s. The mechanism is this: when an individual filament is capped, the retrograde flow makes it retract — the filament would eventually disappear if uncapping does not occur quickly enough. If the filament number becomes less than a minimum needed to overcome the membrane force, the filopodium collapses.

These large-amplitude oscillations of filopodial length observed in our simulations are the consequence of the amplification of molecular noise of capping protein. This amplification

is possible due to the timescale separation between fast polymerization and retrograde flow processes and slow off-rates of formin and capping protein. Since the regulatory proteins are present in very low concentrations (below 100 nM), their noise is highly discrete [65–73], randomly driving back-and-forth transitions from fast growth to fast retraction for each filament. Such a highly-fluctuating behavior should be advantageous from the point of view of efficiency of a filipodia as an environmental sensor, in analogy to near-critical systems, where large fluctuations are indicative of large corresponding susceptibilities. Indeed, we found that the filopodial length is significantly more sensitive to the change in membrane force in a system with capping proteins and formins than in a system with just actin.

In order to compare our predictions with prior and future experiments, we computed filopodial lifetime distribution from 2048 Gillespie trajectories (Fig. 2.2, lower panel). Experimentally reported lifetimes are on the order of several minutes, consistent with our results [60, 74–78]. We found that computed filopodial lifetimes strongly depend on the amplitude of the individual filament length fluctuations (see Fig. 2.3).

To gain further insight into the observed noise amplification phenomenon, we investigated how the capping protein and formin concentrations influence the magnitude of the filopodial macroscopic fluctuations. In order to measure them, we created a stationary state by allowing the tips of disappearing filaments to stay at the filopodial base and eventually uncap. As one might have anticipated, the fluctuation amplitude grows with increasing capping protein concentration (data not shown). On the other hand, formin quenches these oscillations such that the fluctuation amplitude depends mainly on the ratio of the capping protein to formin concentrations (Fig. 2.3, upper panel). Overall, filopodial lifetime strongly depends on the amplitude of the individual filament length fluctuations (Fig. 2.3, lower panel).

To shed light into the kinetics of filopodial retraction, we address the question of how such long timescale processes (hundreds of seconds) emerge from the much faster constituent kinetic rates (filament uncapping rate, $k_C^- = (25\,\mathrm{s})^{-1}$ is the slowest rate in our model). We elab-

orate below on a simple mean-field model to estimate the rate for filament disappearance. We consider a filament in the bundle of N anticapped filaments of stationary length, L, computed in our prior work [15] (see Table 2.1 for the definition of variables):

$$L = \frac{k_D l_D}{N} \left(\frac{C_A \delta}{v_{\text{retr}}} - \left(1 + \frac{k_A^- \delta}{v_{\text{retr}}} \right) \frac{1}{k_{FA}^+} e^{f\delta/N k_{\text{B}} T} \right). \tag{2.1}$$

We keep track of a single filament as various binding/unbinding events occur. After formin unbinds (with off-rate, k_F^-), the filament may either become capped and start retracting or re-bind formin, with the latter being more likely. The ratio of probabilities for capping and anticapping is proportional to the ratio of the on-rates for capping protein and formin, $C_C k_C^+$ and $C_F k_F^+$, where C_C and C_F are capping protein and formin concentrations, respectively. The regulatory proteins are not consumed during polymerization (unlike actin), thus, we can assume bulk concentrations at the tip. Overall, the average capping time for a filament may be estimated as

$$\bar{\tau}_c = (k_F^-)^{-1} (C_C k_C^+)^{-1} (C_F k_F^+). \tag{2.2}$$

The computational model for filopodial dynamics is essentially a multidimensional lattice on which a stochastic propagator enacts a random walk. If the filament is capped, it may either fully retract and disappear or become uncapped and grow back. Thus, many possible trajectories on the lattice must be considered. Each distinct trajectory is characterized by a certain probability of occurrence and the overall time for filament disappearance. The trajectories may be grouped according to how may times the filament has been capped and uncapped before it disappears. In the following analytical estimation, we calculate average time of disappearance in each group of paths described above, using a mean field approximation, instead of carrying out a full path integral calculation (Fig. 2.4). Such a group of paths may be thought of as a particular event scenario. We then average the rates from each group with statistical weights of

the scenarios; the weights are exact.

In a bundle of N filaments the average time to wait for a capping event is $\bar{\tau}_c/N$. Once a filament is capped, it starts to retract under retrograde flow. It may either fully shrink and disappear or uncap and regrow (first trajectory bifurcation, see Fig. 2.4). If uncapping does not occur, it takes $\tau_d \approx L/v_{\text{retr}}$ for the filament to fully shrink and disappear. The probability to follow this scenario is therefore $p = \exp(-k_C^- \tau_d)$. The alternative scenario – the filament uncapping and resuming growth after average time $\bar{\tau}_s$ of shrinking – occurs with $1 - p$ probability. To find the typical shrinking time (the time it takes for a capped filament to uncap), $\bar{\tau}_s$, one has to average over the exponential distribution for uncapping up to τ_d (since for longer times the filament has fully retracted according to the first scenario),

$$\bar{\tau}_s = \int_0^{\tau_d} \tau P(\tau)\, d\tau = \frac{1}{k_C^-} \left[1 - e^{-k_C^- \tau_d} \left(1 + k_C^- \tau_d \right) \right]. \tag{2.3}$$

After uncapping, the filament starts to regrow until the next capping. If x is the position of the barbed end with respect to filopodial base, then the regrowth speed is

$$v_g(x) = \dot{x} = C_A(1 - x/L)k_{FA}^+ \delta - v_{\text{retr}}, \tag{2.4}$$

where the steady-state G-actin concentration gradient has been taken into account (C_A is the bulk concentration at the filopodial base). We solve this equation to obtain the filament length after regrowing for time $\bar{\tau}_c$, the average time until the next capping event:

$$x(\bar{\tau}_c) = \tilde{L} - \left[\tilde{L} - x(0) \right] \exp(-C_A k_{FA}^+ \delta \bar{\tau}_c / L), \tag{2.5}$$

where $\tilde{L} = L \left(1 - \frac{v_{\text{retr}}}{C_A k_{FA}^+ \delta} \right)$. With a very large polymerization rate provided by formin, the exponential factor is essentially zero, so the new length is $x(\bar{\tau}_c) \approx \tilde{L} \approx L$, as $v_{\text{retr}} \ll C_A k_{FA}^+ \delta$. Therefore, if a capped filament is uncapped, it quickly catches up with the others at steady state length. Thus, a filament either retracts after average time $\bar{\tau}_c/N + \tau_d$ with probability p, or

returns to initial state after time $\bar{\tau}_c/N + \bar{\tau}_s$ with probability $1 - p$. This filament will be capped again after average time $\bar{\tau}_c$, and there will be second trajectory bifurcation: it will either retract with probability $(1 - p)p$ at time $\bar{\tau}_c/N + \bar{\tau}_s + \bar{\tau}_c$, or start regrowing again with probability $(1 - p)^2$ at time $\bar{\tau}_c/N + \bar{\tau}_s + \bar{\tau}_c + \bar{\tau}_s$. Each subsequent trajectory bifurcation leads to a longer time (lower rate) for filament disappearance, but also is less likely to occur. To get the full rate one has to average the disappearance rates over all scenarios:

$$\lambda = \sum_{n=0}^{\infty} \frac{p_n}{\tau_n} = \sum_{n=0}^{\infty} \frac{p(1-p)^n}{a+bn} = \frac{p}{a} {}_2F_1\left(1, \frac{a}{b}; \frac{a}{b} + 1; 1 - p\right), \tag{2.6}$$

where ${}_2F_1$ is the hypergeometric function; $a = \bar{\tau}_c/N + \tau_d$; $b = \bar{\tau}_c + \bar{\tau}_s$.

Thus, the average time for filament disappearance is $\bar{\tau} = 1/\lambda$, where λ is given in Equation 2.6. We ran simulations with different numbers of filaments at two formin concentrations, looking for an average time of disappearance of one filament. Since there are multiple filaments in these simulations, and several of them may be capped at the same time, we need to take into account a possibility that the first capped filament is not the first to disappear, in order to directly compare the results from simulations with the analytical estimates. This possibility might occur if the first capped filament uncaps and grows back, while another capped filament quickly retracts at the same time. In Fig. 2.5, the average filament disappearance time is computed by considering not only $\bar{\tau} = 1/\lambda$, but also taking into account the contributions from the simultaneous retraction of other filaments. The comparison between simulations and analytical results, where the latter did not contain adjustable parameters, showed good agreement (see Fig. 2.5).

Another interesting result that we observed was that the longest filopodia grow to about 4.5 times the size of those simulated without formins and capping proteins (in our prior work [15]). The explanation for this follows: a key factor limiting the filopodial length is insufficient actin transport [15, 62]. A mean field estimation for the stationary length (Eq. 2.1) follows from

equating the actin flux from the established gradient and actin consumption by N polymerizing barbed ends [15]. Since the latter is proportional to N, a smaller number of filaments require less actin for growth, therefore, a thinner filopodium can grow longer before the diffusive G-actin transport again becomes a limiting factor. As filaments cap and disappear, N decreases, hence the increase in the length. When too few filaments remain to oppose the force of the membrane load, the bundle withdraws, resulting in the eventual filopodial disappearance. One can speculate another interesting consequence of bundle thinning: it may induce mechanical buckling instability, due to diminution of the mechanical rigidity of the F-actin bundle. In light of the recent suggestion that filopodia with more than 10 filaments are mechanically stable [79], our current finding of filament bundle thinning may turn out to be an important mechanism for mechanical collapse of filopodia due to buckling.

In summary, we have shown in this work that capping proteins exert a dramatic effect on filopodial dynamics. Introducing them allowed us to create the first computational model which predicts a finite lifetime of filopodia. The resulting filopodial length dynamics has a remarkable feature: discrete noise of regulatory proteins, that are in very low concentration, becomes greatly amplified. Because of timescale separation, this slow discrete noise triggers the fast retrograde flow and polymerization processes resulting in oscillations of filopodial length macroscopic in space and time. This result suggests that experimentally observed filopodial growth-retraction turnover dynamics [59–61] may be partially driven by the internal noise of the filopodial mechano-chemical network. Large amplitude fluctuations in filopodia may be important with respect to their sensory role. In particular, a molecular system having oscillations of this magnitude is expected to be easily perturbed by small external forces, either mechanical or chemical in origin, in analogy with large susceptibilities seen in near-critical systems with large fluctuations. The noise amplification described in this work, arising from discrete noise in a low copy number of some of the reaction network species, is related to stochastic switching in biochemical signal transduction, for example when a cell needs to

make a binary decision [80]. A similar dynamic instability is observed in microtubular growth, although in filopodia the oscillations are switched by the binding noise, and in microtubular catastrophes and rescues it is enzymatic in origin [81]. In addition, by promoting filament disappearance, capping proteins thin the filopodium, reducing its mechanical rigidity. We speculate that this may be an important mechanism for creating buckling instabilities. We also investigated in detail the filament retraction kinetics, and developed a mean field analytical estimation for the rate of disappearance of individual filaments that matches the simulation data surprisingly well. We explained the extremely slow timescale of filament disappearance, compared to bare kinetic rates, by common occurrence of multiple capping, shrinking and regrowing events, before the filament fully retracts as a result of a rare fluctuation.

In our stochastic model of filopodial growth [15], represented in Fig. 2.1, we assume that the reaction processes are confined to a spatial region having a linear dimension of ζ (so-called Kuramoto length [82]), such that particles diffuse across the region quickly compared to the typical reaction times. This allows us to discretize space into compartments and model protein diffusion along the filopodium as a random walk on a one-dimensional lattice, with molecules hopping between these compartments at rates calculated from diffusion coefficients. At relevant concentrations, ζ is on the order of 100 nm at the tip of the filopodium. We chose a somewhat more conservative compartment size of $l_D = 50$ nm [15]. At the filopodial base the protein concentrations are kept at their constant bulk values. Biochemical reactions within each compartment are simulated using the Gillespie algorithm [15, 83]. Polymerization rates at the tip are decreased by the membrane force via the Brownian ratchet model [15, 62]. The experimentally reported uncapping frequency is about once every 30 min, however, it may be greatly increased through the actions of uncapping proteins, such as PIP2 [64]. The variables and parameters for the model are given in Table 2.1. Further simulation details are elaborated below.

2.4 The stochastic model of filopodial growth and retraction under the influence of capping proteins

A mature filopodium is a bundle of a few dozen actin filaments enclosed by the cell membrane with a protein complex at the tip [84]. The filopodial base is usually located within a three-dimensional actin mesh below, which forms the basis for the lamellipodium [84]. In addition, bundling proteins cross-link the filaments in the bundle, however, this was not included in our current model.

Typical growth and retraction speeds are within $0.1 - 0.2\,\mu\text{m/s}$, with filopodial lengths reaching around $1 - 2\,\mu\text{m}$ [85]. In special cases, the filopodial length may reach nearly one hundred microns [86]. Since short filopodia are mostly straight, for simplicity we assume growth of straight filaments. Longer filopodia might tilt and bend. The possibilities of bending of the filaments and buckling of a filopodium will be addressed in future work.

Depending on the number of actin filaments, the filopodial diameter typically varies between $d = 100$ nm and 300 nm. In our simulations we used $d = 150$ nm, a number which was derived from minimizing the membrane free energy [63] and is reasonable if the number of filaments is not too large.

Our physico-chemical model of filopodial growth consists of the following processes: 1) diffusion of proteins along the filopodial tube, providing passive transport from cell body to the filopodial tip; 2) mechanical interactions between the membrane and individual filaments; 3) actin filament polymerization and depolymerization processes at the barbed end; 4) depolymerization at the pointed end and the induced retrograde flow v_{retr} of the bundle as a whole; 5) capping of the filaments that stops polymerization [64]; 6) binding of formin to the barbed end which increases the effective polymerization rate fivefold [10]

2.4.1 Polymerization, depolymerization

Actin filaments (F-actin) are asymmetric, with one end called "the barbed end" and the other end called "the pointed end". Due to ATP binding and subsequent hydrolysis, chemical affinities for monomers (G-actin) are different at the two ends, such that the polymerization rate at the barbed end is much higher. This leads to the motion of the filament as a whole in the direction of the barbed end (while individual monomer units migrate from the barbed end to the pointed end). This process is called "treadmilling" and is the biochemical basis for the cytoskeletal dynamics [87, 88], including filopodial protrusion. An actin filament consists of two protofilaments, wound up in a right-handed helix, with a pitch distance of 37 nm. Diameter of one globular actin monomer is 5.4 nm. In our simulations we increase the filament length by $\delta = 2.7$ nm upon one polymerization event, since there are two protofilaments, and two monomers are needed to increase length by 5.4 nm.

The resulting double helical filament is mechanically robust, with a persistent length, $L_p \sim$ 10 μm. The buckling length, or the critical length at which one filament would buckle if subject to a force F, is

$$L_b \approx \frac{\pi}{2}\sqrt{\frac{k_{\mathrm{B}}TL_p}{F}}, \tag{2.7}$$

where k_{B} is the Boltzmann constant and $k_{\mathrm{B}}T = 4.1$ pN nm at room temperature [62]. The equation gives $L_b \approx 100$ nm for a force $F = 10$ pN for a single filament. For weakly cross-linked bundles of N actin filaments, the buckling length of the bundle is $\sqrt{N}L_b$, while it is $NL_b/\sqrt{2}$ for tightly linked bundles [62]. A recent work suggested that the membrane enclosure significantly increases these estimates [79].

In our model, cytosolic molecules such as G-actin, formin or capping protein, that are in the same compartment as the filopodial tip can attach to one of the filament ends with a probability, given by the rates k_A^+, k_{FA}^+ or k_C^+. These monomers at filament ends can also stochastically

dissociate with the corresponding rates k_A^-, k_{FA}^- or k_C^-. In prior experiments, it was shown that the actin polymerization rate rather strongly depends on the membrane force [89]. Theoretical explanations were also provided [62, 90, 91]. For instance, the Brownian ratchet model considers membrane fluctuations at the tip of the filopodium [90, 92]. If the membrane fluctuations are sufficient to allow a G-actin monomer to fit sterically atop the filament, that allows a polymerization with the rate k_A^+. Thus, the effective polymerization rate $k_{A,n}$ on the n-th filament equals the "bare" rate $k_{A,0}$ times the probability of the gap opening at the tip of the n-th filament. A convenient relation between the loading force f_n and the polymerization rate was derived earlier [92],

$$k_n^+ = k^+ \exp(-\frac{f_n \delta}{k_B T}). \qquad (2.8)$$

According to Eqn. 2.8 the on-rates for G-actin monomer addition (k_A^+ for free barbed end; k_{FA}^+ for the barbed end anticapped by formin) and capping protein (k_C^+) are modified after each timestep. The on-rate for formin, k_F^+, was assumed to be independent of the membrane fluctuations due to diminished steric constraints [10]. All of the off-rates do not depend on the membrane position.

For these effective on-rates, we needed to estimate the membrane force on each individual filament, which is discussed next.

2.4.2 Membrane force

According to prior studies fluctuations of a membrane sheet below micron scale relax on the microsecond to millisecond timescale [93–96]. Chemical reactions occur on a much slower timescale ($10^{-2} - 10^1$ s), thus, due to this timescale separation, the membrane fluctuations may be assumed to be equilibrated at the time of each given chemical event. Each filament experiences an individual membrane force, f_n, that depends on the proximity of the filament

tip to the average location of the membrane,. Thus, the total membrane force f is distributed among the individual filaments, $f = \sum_n f_n$.

In order to calculate f_n, we assume that on average the force on a filament is proportional to the membrane-filament contact dwelling probability, which is the probability that the membrane touches that filament. That, in turn, depends on the amplitude of the membrane fluctuation near the filament and the filament length. Longer filaments are more likely to be in contact with the membrane, and feel a stronger membrane force. If the membrane fluctuations are assumed to be described by a Gaussian distribution around the membrane average position, the dwelling probability p_n for the n-th filament to be in contact with the membrane is proportional to the probability that the membrane height is found below the filament end

$$p_n \sim \int\limits_{h-h_n}^{\infty} \exp(-z^2/\sigma_d^2)\,dz, \tag{2.9}$$

where σ_d is the average membrane fluctuation amplitude (discussed next). Once p_n is obtained, the force f_n on each filament may be computed,

$$f_n = \frac{p_n}{p}f, \tag{2.10}$$

where $p = \sum_n p_n$ is a normalization factor. Then we use Eqn. 2.8 to modify the polymerization rates.

2.4.3 Retrograde flow

The filopodial dynamics are strongly controlled by the polymerization at its tip [7, 97]. As was mentioned earlier, the "treadmilling" contributes to a steady backward motion of the whole actin filament bundle, called the retrograde flow. In some cells, it is believed that specific myosin motors participate in creating the retrograde flow [12]. All of this is subject to the

regulation by the signaling proteins, but in our current computations, we neglect these subtleties of the retrograde flow process, and assume a constant average retrograde flow speed, v_{retr}. The retrograde flow speed v_{retr} is approzimately $20 \sim 200$ nm/s [97–101] and we take $v_{retr} = 70$ nm/s as the default value in our simulations.

2.4.4   Numerical Scheme

Our computations are based on the Gillespie algorithm [83]. In particular, at each simulation step two independent random numbers are chosen, where the first one determines the time of the next event $t + \Delta t$ (t is the time of last event) based on the reaction rates present in the system, and the second one determines which event occurs, based on the magnitude of individual event rates. The event is chosen among the following possibilities: a) G-actin monomer, capping protein or formin hopping between filopodial compartments, b) individual filament polymerization or capping events, c) depolymerization or uncapping events and d) binding or unbinding of formin. Then we update current time, adding Δt, the numbers of proteins in compartments, filament length and which protein is on its end according to the event that happened. We then incorporate retrograde flow and update the length of filaments again. Finally, the membrane load is partitioned among the filaments, as described above, which results in recomputing of individual filament polymerization rates, according to the Brownian ratchet model.

Table 2.1: Model variables and parameters in the simulations with capping proteins.

Mechanics	
Half actin monomer size	$\delta = 2.7$ nm
Number of filaments	$N = 16$
Membrane force	$f = 10$ pN
Diffusion rate	$k_D = 5\,\mu\text{m}^2\text{s}^{-1}(2000\,\text{s}^{-1})^*$
Membrane fluctuation	$\sigma_d = 20$ nm
Retrograde flow speed	$v_{\text{retr}} = 70$ nm/s

(De)Polymerization rates	
Free barbed end	$k_A^+ = 11.6\,\mu\text{M}^{-1}\text{s}^{-1}(21.8\,\text{s}^{-1})$
Formin-anticapped	$k_{FA}^+ = 53.2\,\mu\text{M}^{-1}\text{s}^{-1}(100\,\text{s}^{-1})$
Depolymerization	$k_A^- = 1.4\,\text{s}^{-1}$

(Anti)Capping rates	
Formin on-rate	$k_F^+ = 10\,\mu\text{M}^{-1}\text{s}^{-1}(18.8\,\text{s}^{-1})$
Formin off-rate	$k_F^- = 0.0667\,\text{s}^{-1}$
Capping rate	$k_C^+ = 3.5\,\mu\text{M}^{-1}\text{s}^{-1}(6.6\,\text{s}^{-1})$
Uncapping rate	$k_C^- = 0.04\,\text{s}^{-1}$

Bulk concentrations	
Actin	$C_A = 10\,\mu\text{M}$
Capping protein	$C_C = 50$ nM
Formin	$C_F = 40$ nM; 80 nM

[*] Reaction and diffusion rates in parenthesis in "seconds" units depend on the compartment volume. In our computations it was fixed, with compartment length of $l_D = 50$ nm and filopodial diameter of 150 nm.
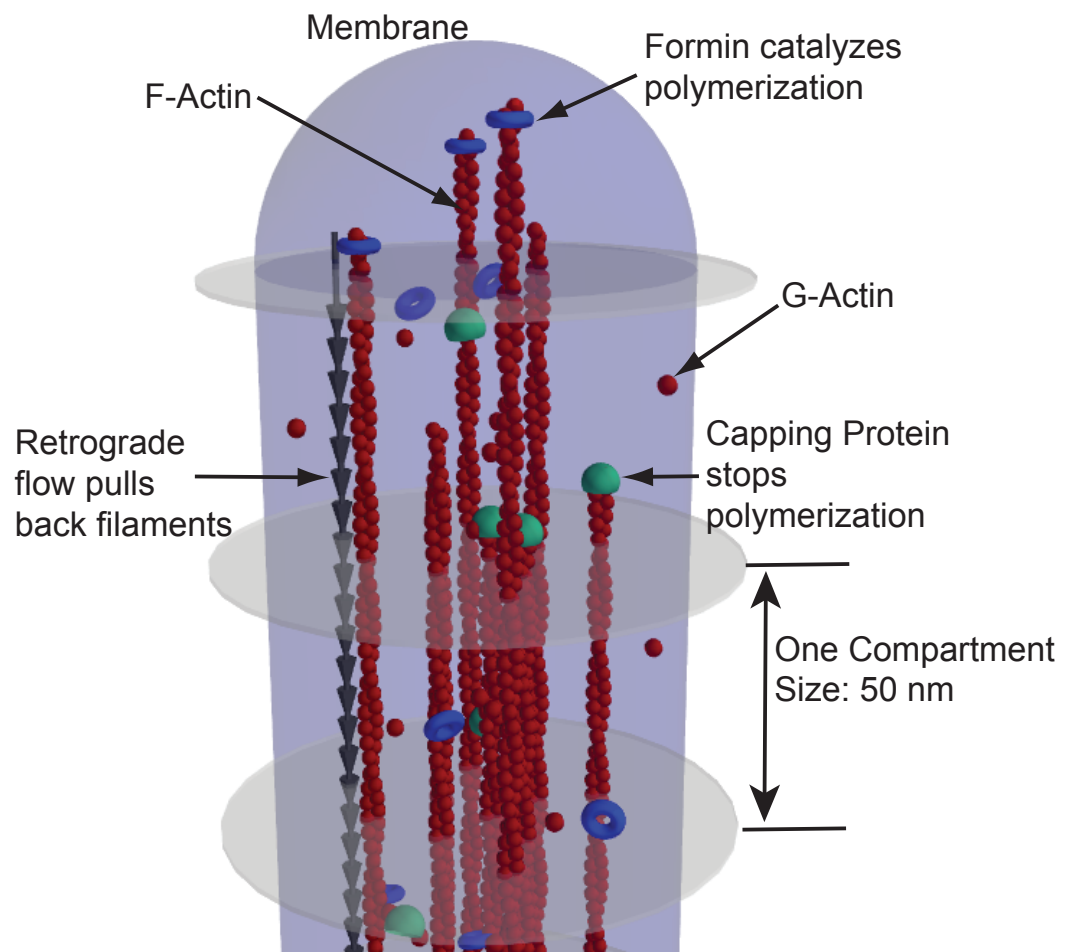
Figure 2.1: **The mechano-chemical model of the filopodium.**
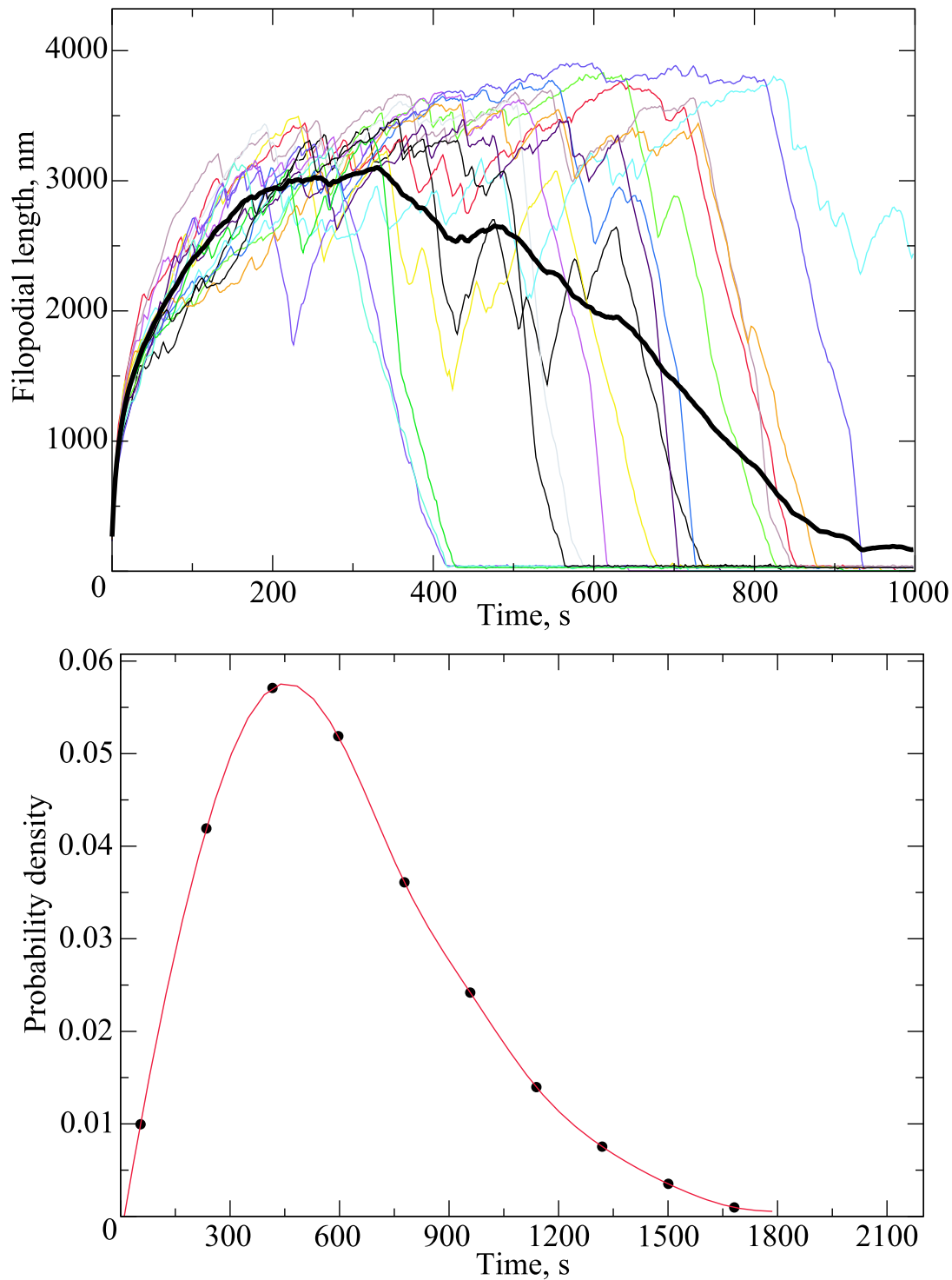
Figure 2.2: **Filopodial growth and lifetime distribution.**
Upper panel: 16 individual trajectories are shown from stochastic simulations of filopodial growth and retraction with 80 nM of formin and 50 nM of capping protein. The average over trajectories is indicated with a thick black line. Individual trajectories undergo turnover – growth-retraction oscillations on a micron scale – which is induced by molecular noise of regulatory proteins. *Lower panel*: Distribution of the filopodial lifetime calculated from 2048 trajectories at 40 nM of formin and 50 nM of capping protein.
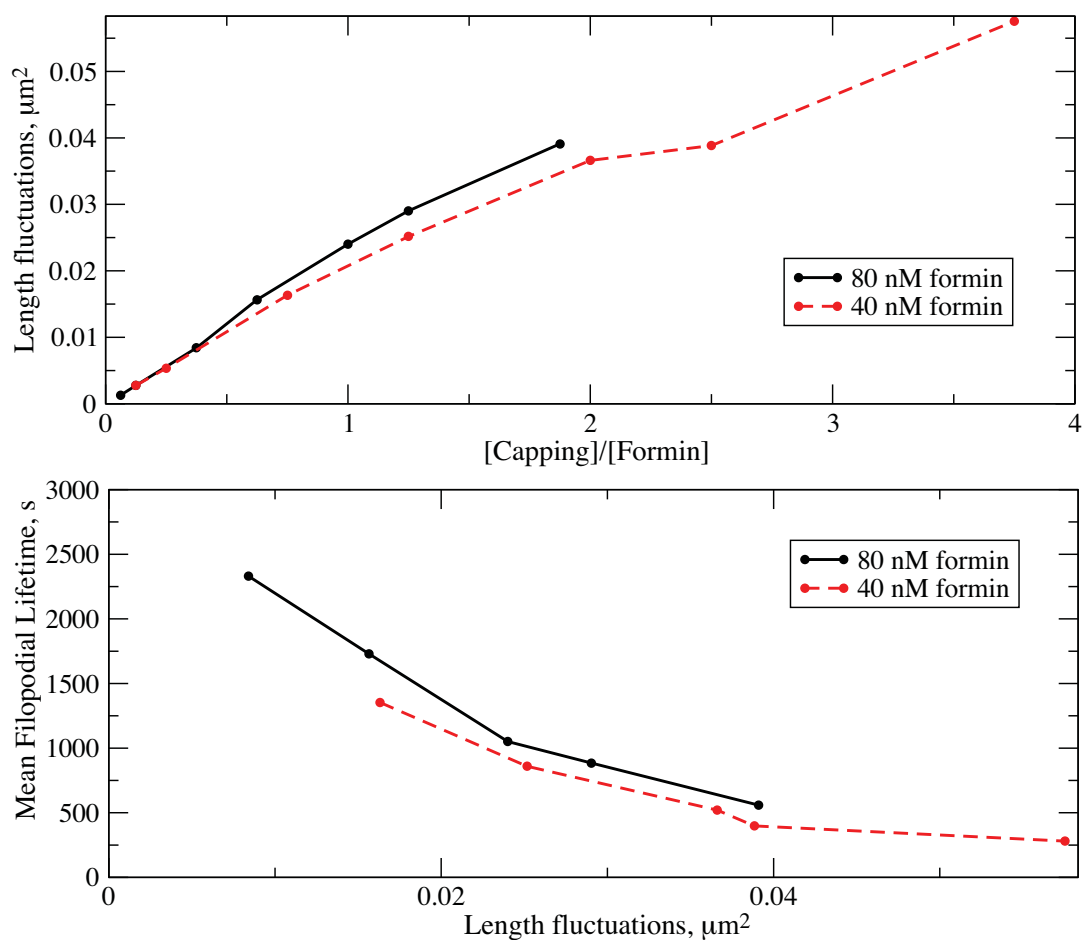
Figure 2.3: **Filopodial length fluctuations**

The amplitude of filament length fluctuations depends on the concentrations of regulating proteins, mainly, on their ratio. These fluctuations, in turn, largely determine the mean filopodial lifetime.
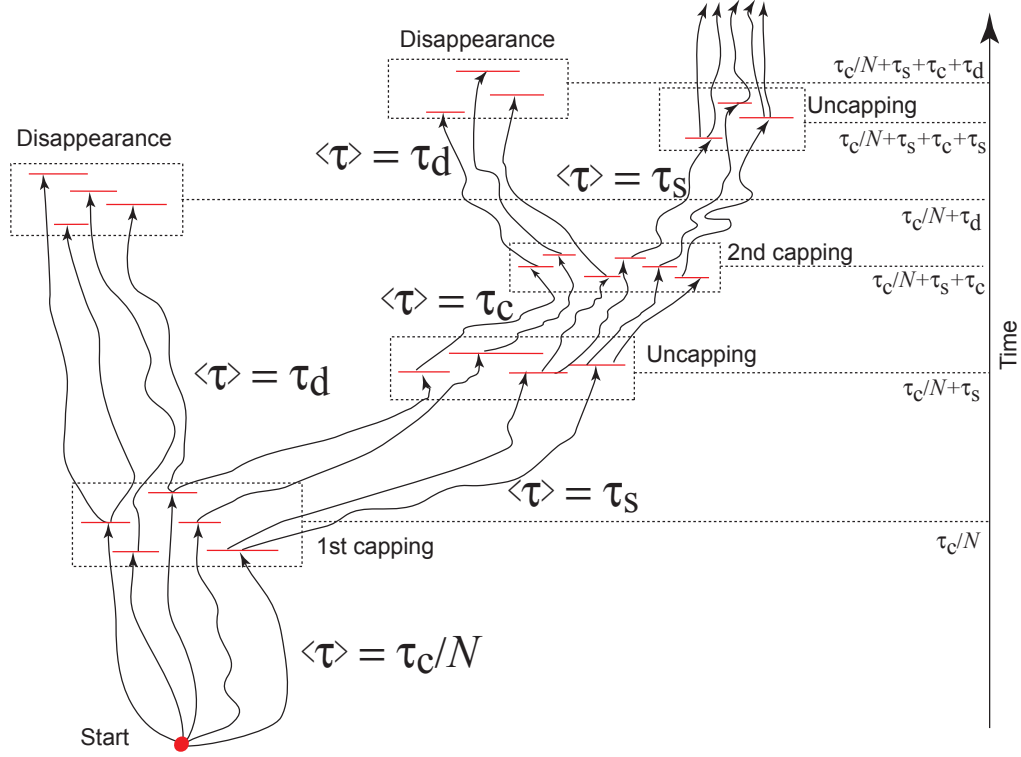
Figure 2.4: **Grouping of the stochastic trajectories into different capping-uncapping scenarios.**
The system dynamics is executed on a multidimensional lattice on which a stochastic propagator enacts a random walk. Eventually, a trajectory arrives to a sublattice where one filament is capped. Mean-field average time for that is $\bar{\tau}_c/N$. From there some trajectories lead to filament disappearance due to retrograde flow and some to filament uncapping with respective statistical weights of p and $1-p$ and mean field average times τ_d and $\bar{\tau}_s$. After uncapping, all trajectories will pass through the phase area where this filament is capped again after mean field average time $\bar{\tau}_c$. Then again, it may either disappear after time τ_d or uncap after time $\bar{\tau}_s$ with weights p and $1-p$.

34
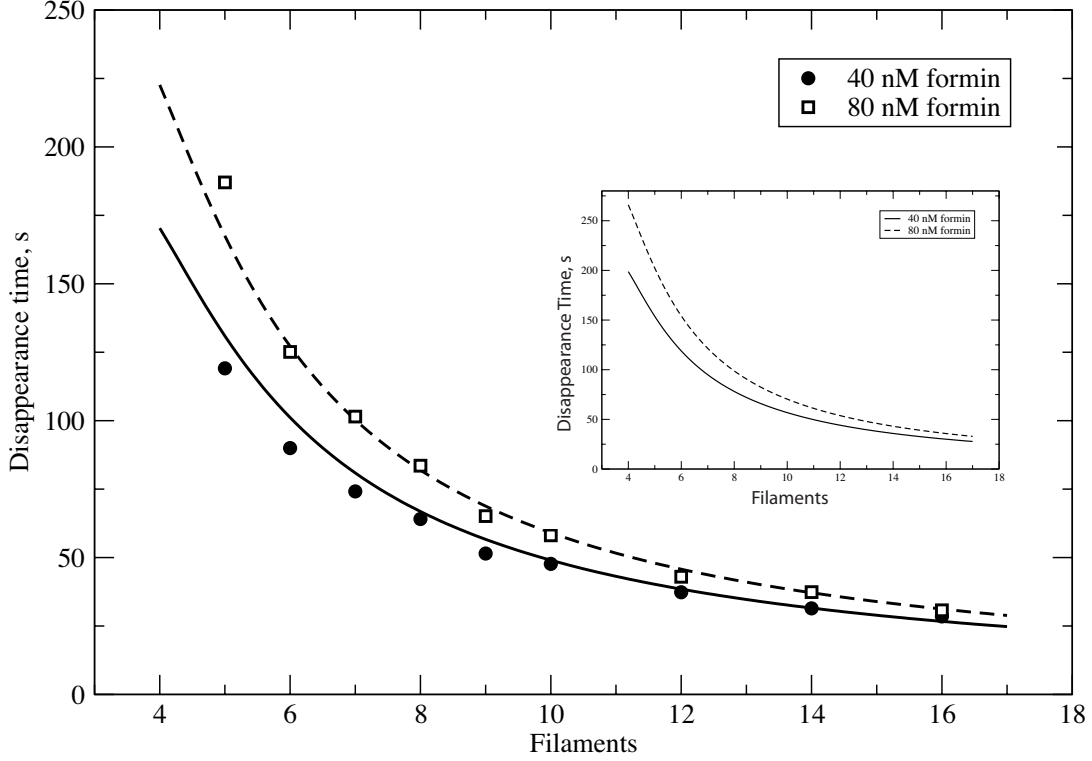
Figure 2.5: **Rate of individual filament disappearance from a bundle of N filaments.**
C_C=50nM and C_F=40nM (circles, solid line) and 80nM (squares, dashed line). The symbols
show simulation results. Solid and dashed lines are computed from combining the disappear-
ance for a single filament, given by Eq. (2.6), with the scenarios of other filaments disappearing,
which had capped later. We used the following expression to estimate the average time for any
of the bundle filaments to fully retract, $\tau_{\text{multi}} = (1-p)^{n_c-1}\lambda^{-1} + (1-(1-p)^{n_c-1})(\bar{\tau}_c n_c/N +
\tau_d)$, where, $n_c = \lambda^{-1}/(\bar{\tau}_c/N)$, indicates the average number of simultaneously capped fila-
ments, while the remaining symbols are defined in text. In this expression, the first term cor-
responds to the first capped filament disappearing first, while the second term accounts for the
possibility of one of the other $n_c - 1$ capped filaments following the scenario of quick disap-
pearing without uncapping even once. The inset shows the bare disappearance times for single
filaments, λ^{-1}, computed from Eq. (2.6), where the effect of other filament retractions is not
taken into account.

Chapter 3

Design of Active Transport Must Be Highly Intricate: a Possible Role of Myosin and Ena/Vasp for G-Actin Transport in Filopodia

Pavel I. Zhuravlev[†], Bryan S. Der[‡], Garegin A. Papoian[†]

[†]*Department of Chemistry, University of North Carolina at Chapel Hill, North Carolina, 27599*

[‡]*Molecular and Cellular Biophysics Program, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599*

3.1 Abstract

Recent modeling of filopodia – the actin-based cell organelles employed for sensing and motility – reveals that one of the key limiting factors of filopodial length is diffusional transport of G-actin monomers to the polymerizing barbed ends. We have explored the possibility of active transport of G-actin by myosin motors, which would be an expected biological response to overcome the limitation of a diffusion based process. We found that in a straightforward implementation of active transport the increase in length was unimpressive, not more than 30 per cent, due to sequestering of G-actin by freely diffusing motors. However, artificially removing motor sequestration reactions led to about 3-fold increases in filopodial length, with the transport being mainly limited by the motors failing to detach from the filaments near the tip, clogging the cooperative conveyer belt dynamics. Making motors sterically transparent led to a qualitative change of the dynamics to a different regime of steady growth without a stationary length. Having identified sequestration and clogging as ubiquitous constraints to motor driven transport, we devised and tested a speculative means to sidestep these limitations in filopodia by employing cross-linking and putative scaffolding roles of Ena/VASP proteins. We conclude that a naïve design of molecular motor based active transport would almost always be inefficient — an intricately organized kinetic scheme, with finely tuned rate constants, is required to achieve high-flux transport.

3.2 Introduction

For processes including cancer metastasis [6], neuronal growth [4], wound healing [3], and embryonic development [102], cell motility is guided by the sensing function of finger-like projections called filopodia. The final location of the cells is critical, so these cells must adeptly sense their environment to properly direct their movement. To shed light into these processes, the underlying physical and regulatory mechanisms of filopodial growth and retraction must

be understood. In fact, the overall structure and function of filopodia is now largely known: G-actin monomers polymerize into F-actin filaments, which then bundle in parallel to protrude the cell membrane. Continued polymerization of G-actin at the barbed ends and depolymerization at the pointed ends results in "treadmilling", one of the key processes in the dynamics of actin-based cellular structures [87, 88].

In treadmilling, F-actin elongation results from polymerization at the barbed ends, while retraction results from depolymerization at the pointed ends, backward pushing from the stretched membrane, and ATP-driven pulling from the cell body (the latter two processes are termed retrograde flow [103, 104]). Polymerization and retrograde flow are fast processes, and switching their equilibrium engenders complex and highly dynamical behavior, such as growth-retraction cycles and turnover process [2,59,61,105,106]. These dynamics allow a filopodium to perform its role as a mechanochemical receptor to guide cell motility.

A more detailed understanding of filopodia, beyond this general picture of overall structure and function is a challenging pursuit because of the complex interdependence of governing mechanical, chemical, and biological processes. Thus, the development of computational models which would produce quantitative and testable predictions can be an effective means toward gaining additional new insights.

Considerable efforts have been made in modeling filopodia and other organelles based on polymerizing bundles of F-actin, including stereocilia and microvilli [15,62,63,105,107–109]. These models, some deterministic, some stochastic, suggest that the diffusional flux of G-actin to the polymerizing end, in the absence of other chemical or mechanical regulation, is a limiting factor determining the filopodial length [15, 62]. While experimentally measured filopodia are often several microns in length, some reach 30, or even 100 μm [60, 86, 110]. Interestingly, physiologically reasonable choices of parameters in the current computational models predict filopodial lengths which are many-fold shorter than longest filopodia observed experimentally. In addition, the growth rates of the longest filopodia are on the order of 10 $\mu m/min$ [60]. This

value is also many-fold higher than current models have reported. These discrepancies indicate that a fundamental mechanism of growth is currently unaccounted for. For this reason, we extended our previous model [15, 105] to investigate how a hypothetical active transport of G-actin might influence filopodial length. ATP-driven molecular motors are employed for many transporting purposes in cells [111], providing a directed and faster alternative to diffusion. Will active transport of G-actin in filopodia sufficiently promote elongation to fill the gap between models and experimental observations?

This hypothesis of active transport goes against the grain of the long-held view of passive diffusion in the context of actin filament growth [112]. Although polymerization is clearly diffusion-limited *in vitro* [113], an *in vitro* system necessarily omits much of the *in vivo* complexity, and mechanisms may vary across actin-based structures. In fact, an *in vivo* study tracking G-actin within a lamellipodium showed that G-actin migration is too fast to be explained by passive diffusion alone [114].

Molecular motors are a likely candidate for supplementing the passive diffusion of G-actin during filopodial elongation. Myosins, the family of proteins that can "walk" on actin filaments, have been detected in both filopodia and stereocilia. In fact, the importance of myosin X(M10) [115, 116] in filopodial growth is well-supported experimentally [104, 117], though the specific roles and cargo of various myosins remain unclear. As new myosin motors and their functions are still being identified [16], it may be valuable to use computational models to generate testable hypotheses for the role of these motors.

Myosin X has been identified at the filopodial tips [16], and its overexpression resulted in increased number and length of filopodia in motile cells [16]. Furthermore, myosin X has been observed walking forward and moving rearward within filopodia [118]. The forward motion suggests directed movement of motors along a filament that significantly outraces retrograde flow, and the slower rearward movement suggests stochastic periods without walking and backward movement due to retrograde flow.

Beyond the identification of such motors within filopodia, there are a variety of hypotheses for their actual role in filopodial growth – while not mutually exclusive, they include a physical pushing against the membrane to effectively increase the polymerization rate [117], transport of integrin to form adhesive structures near the tip [119], and transport of regulatory proteins, such as Ena/VASP family members, toward the tip [17]. The complete role of Ena/VASP in filopodial growth is not fully understood [11, 120, 121] though it has both G-actin and F-actin binding motifs [122], forms a tetramer to crosslink actin filaments [120–123], and is transported by myosin X [17].

We explore the hypothesis that myosin motors carry G-actin to the polymerizing barbed ends of the actin filaments, presumably widening the main bottleneck for the filopodial length. To our knowledge, only one experimental study suggested a mechanism of actin transport that goes beyond passive diffusion [114]. In addition, a theoretical study, based on mean-field deterministic equations, investigated the distribution of various proteins along a stereocilium under the influence of motor proteins [107]. The current paper represents the first filopodial model of active transport, building upon our previous work [15, 105] to stochastically simulate individual motors walking on actin filaments and carrying G-actin inside filopodia.

The stationary length in our previous model is set by a balance of fluxes: G-actin diffuses forward to the tip along its gradient; all of that flux is consumed for polymerization to F-actin; and as the F-actin filaments are pulled back by retrograde flow, an equal flux but opposite in direction returns actin to the cell bulk. Transport by motors adds to the forward flux: this addition is directional, as in retrograde flow, and not linearly decreasing with length like diffusional flux [15, 62, 107]. If motor-based flux dominates the diffusional flux, then motor-based transport of actin indeed may be visualized, similar to cartoons in biology textbooks, as a conveyor belt. Furthermore, if active transport flux exceeds the retrograde flow flux, there will be no stationary length.

Despite this possibility, straightforward introduction of motors did not lead to skyrocketing

of the stationary filopodial length, let alone indefinite filopodial growth. In fact, freely diffusing motor proteins sequester G-actin previously available for polymerization. As a result, motors at large concentrations play a length-diminishing rather than a length-promoting role. In cases when simulated filopodia did grow longer, it was only a modest increase in length. Therefore, sequestration by motors can severely undermine active transport, and should be avoided to achieve a noticeable effect.

We repeated the simulations while disabling sequestration, by forbidding the freely diffusing motors (as opposed to motors bound to and walking along a filament) to load G-actin from the solution. This yielded a 3- to 5-fold increase in filopodial length. Interestingly, we also noticed that the longest filopodia were observed in cases when motors had lower affinity to the filaments. This is a manifestation of "clogging" of the filament "rails" by the empty motors closer to the tip. We ran control simulations where motors loaded with actin are allowed to pass through empty motors: the most common result was large increase in stationary lengths, even in one case resulting in unsaturable linear growth.

Having identified limiting factors to motor-assisted filopodial growth, we explored a plausible biological means to side-step these limitations. In these simulations myosin X does not load G-actin directly, instead loading Ena/VASP as a multi-site adaptor, which in turn binds several G-actins. In this model, Ena/VASP is irreversibly consumed at the tip by a sink mimicking its cross-linking role [11] with simultaneous release of G-actin to the solution. Consequently, sinking of Ena/VASP for cross-linking results in decreased Ena/VASP concentration in solution, diminishing G-actin sequestration. The largest observed stationary lengths in this set were many-fold longer compared with filopodia produced using a "naïve" model of motor transport.

We summarize these findings as *the rules of active transport*: an effective transport mechanism must successfully overcome sequestration of the cargo and clean the "clogged rails". It is likely that these are general principles that apply beyond the context of active transport in filopodia, where they have been deduced.

The results presented in this paper were obtained from about 8000 simulation trajectories, having taken $\sim 100\,000$ CPU hours on UNC Topsail supercomputer.

3.3 Results

The basic components of our filopodia model include polymerization, depolymerization, G-actin diffusion, retrograde flow and membrane force [15]. In the first set of simulations reported here, we added motor molecules that can diffuse, load actin, attach to the filaments, walk on filaments, detach and unload actin. Bulk G-actin concentration influences steady-state length [15], thus it was kept consistent between simulations. Accordingly, we placed corresponding boundary conditions at the bottom of a filopodium (where it emerges from the cell's leading edge), assuming equilibrium with the motor-loading reaction (M10+G-actin) in the cell bulk, such that the sum of freely diffusing G-actin and G-actin loaded on freely diffusing motors was kept constant, at $C_A = 10\mu M$. G-actin binding to motors allows transport but can also result in sequestration, leaving little free G-actin available to polymerization at the filament barbed end if motor-actin affinity is too high. Thus, rate of actin dissociation from motors (k_u) is one key parameter we explore. We also explore the rate of motor dissociation from filaments (k_{fu}) and motor concentration [M].

Exploring the 3-parameter space (k_u, k_{fu}, [M], see Table 3.1 for other parameters) via 8 x 8 x 8 logarithmic grid (that is each parameter was scanned on a logarithmic scale, e.g, [M]=0.1, 0.3, 1, 3, 10, 30, 100, 300 μM), we covered the biologically plausible ranges for these parameters. Intuitively, very low [M] or very high k_u are equivalent to the absence of any motors. In the latter case, motors are present, but do not carry actin. Thus, k_u determines the fraction of motors that do carry G-actin. For this reason, the stationary length depends mostly on the ratio of k_u and [M] (as can be seen by partial collapse of curves in Fig. 3.2): very high motor concentrations or low k_u result in sequestration of G-actin and filopodia do not grow. There-

fore, in many cases the stationary length turned out to be shorter than without motors. In a few cases, we observed about 30% increase in stationary length with the intermediate parameter values (Fig. 3.2). In Fig. 3.3 the 3-parameter space is given in three 2D projections and the bottom-right diagram in Fig. 3.3 includes all 3 parameters. The regions where filopodia do grow longer than they would without motors are marked to provide a phase-diagram-like comprehensive representation. These turn out to be quite localized regions, with low motor concentration. Counterintuitively, upregulating the motors will not lead to increase in transport efficacy because of sequestration. Moreover, motors accumulate at the tip, in qualitative agreement with previous mean-field calculations in stereocilia [107], strongly sequestering G-actin in the location where G-actin's concentration is initially low and where it is most needed for polymerization. In the stationary state diffusional backward flux of the motors due to concentration gradient is equal to the flux of their directed motion forward.

Realizing that sequestration limits length, we ran another set of simulations within the same parameter space, but with actin-loading of diffusing motors (as opposed to filament-bound motors) turned off. Conceptually, motors falling off filaments (accumulating at the tip) did not sequester actin. The goal of these simulations set was to confirm the limiting role of sequestration and to find out what would limit the length in its absence. Also, since binding of G-actin by myosin X has not been observed experimentally, this scheme is not necessarily unrealistic.

The longest filopodia observed in this set of simulations were almost 6 μm, more than 7 times longer than those that grow in a model without active transport. However, the increase could have been merely an effect of increased presence of G-actin in the filopodial "tube", as in addition to those molecules that freely diffuse in cytosol (maintained to be 10 μM at the bottom) some are carried by the motors attached to the filaments. To verify that directed transport was responsible for this length increase, we ran two control simulations: In one we set motor backward step rate $k_{\leftarrow}$ to be equal to forward step rate $k_{\rightarrow} = 50 s^{-1}$ making the motors

44

non-directional random walkers. The resulting filopodial lengths were back to those without motors. Secondly, we set $k_{\leftarrow} = k_{\rightarrow} = 2000 s^{-1}$, similar to diffusion rate, in which case the increase in length was minuscule compared to directional motors even at these unrealistically high motor speeds.

The largest lengths in the nonsequestering motors model were achieved at the k_{fu} values on the higher end of the parameter space (30 s^{-1}, 100 s^{-1}, 300 s^{-1}), corresponding to lower affinity between motors and F-actin. On one hand, high k_{fu} helps to clear the filaments from the motors that have delivered cargo at the tip; on the other hand, if very few motors stay on the filaments the transport is inefficient. Our results show that the former outweighs the latter: the high values of k_{fu} for the longest filopodia indicate that the "clogging" of filaments by empty motors represents yet another key bottleneck for achieving long stationary lengths.

To gauge the influence of this bottleneck, we created an artificial setup where loaded motors could step forward even if the site was occupied by an empty motor (i.e. pass through each other). In one case ($k_u = 3s^{-1}, k_{fu} = 300s^{-1}, [M] = 1\mu M$) such a scheme results in a qualitatively new growing regime. In the new conditions the forward directional actin flux overcame the backward directional flux due to retrograde flow, providing the unquenchable source of actin for the polymerizing barbed ends (see Fig. 3.4 juxtaposing the two regimes). In these simulations filopodia continued to grow linearly for over 100 s up to more than 6 microns, without any indication of growth saturation, where the latter has been found in all other cases discussed here or in prior works [15, 62].

Thus, we have learned that sequestration and clogging can stunt motor-assisted filopodial growth. If active transport is indeed employed in living organisms for G-actin supply of the polymerizing barbed ends, there has to be a biological mechanism for disabling these limitations. In search of a plausible biological mechanism, we turned to Ena/VASP, which has traditionally been implicated in anti-capping activity but actually has additional functions [120]. Ena/VASP can bind G-actin [122, 124], and it has also been observed as myosin X cargo [17].

However, from ideas in our previous works [15, 105], we conclude that a protein only needs to be transported actively if it is consumed during the growth, like G-actin. Regulatory molecules, like capping or anti-capping proteins, that only interact with the filament ends, are required in only very limited amount, and at realistic growth speeds diffusion will always provide sufficient flux. Therefore, we suggest that active transport of Ena/VASP indicates that it may be continuously consumed, most likely to cross-link the filaments [11]. There can be other reasons to be motor-transported. For instance, as we investigate here, Ena/VASP might serve as a scaffold, or an adapter, between myosin X and actin monomers. In this scenario of a "sinking adaptor", motors would not bind and sequester G-actin when the adapter is absent (and consumption due to the cross-linking will diminish adaptor concentration), though clogging would not necessarily be avoided.

To explore this sinking adaptor possibility, we allowed Ena/VASP to bind up to 2 G-actin monomers and/or be loaded to a motor. We simulate a maximum of 2 G-actin molecules transported per motor for simplicity, though one motor could carry up to 8 G-actin molecules: Ena/VASP can form tetramers [122] where each tetramer, therefore, has four G-actin binding sites. Myosin X is also a dimer with two heads and two tails [116,125,126], and each tail could bind an Ena/VASP tetramer as cargo (not considered in our scheme for simplicity). This 8:1 stoichiometry could lead to efficient transport, providing that sequestration is avoided.

In the first set of Ena/VASP simulations, a cross-linking sink was not incorporated. Sequestration still occurred, and filopodia lengths were similar to those with "naïve" motors. In the second set of Ena/VASP simulations, a cross-linking sink was incorporated. Ena/VASP does have a bundling role [11] and is localized near the tip [120] (where sequestration is most deleterious), so it is possibly consumed there for cross-linking, releasing actin for polymerization. Bundling would provide structural rigidity for long filopodia, but because we assume straight filopodia, the desired scenario is simply achieved by reactions of Ena/VASP irreversible sinking near the tip with simultaneous release of any G-actin it had carried.

Some of the simulations with this scheme did show significantly increased stationary lengths compared to simulation of naïve motors. In the most interesting case ($k_u = 30s^{-1}, k_s = 30s^{-1}, k_{fu} = 100s^{-1}, [M] = 1\mu M$), the growth seemed to be linear for over 300 s, though the slope was very gradual. This corresponds to a small difference between retrograde flow actin flux and motor-transported actin flux, with the former slightly larger, so that it dominates until the filopodium becomes very long. The stationary length here reached 3.5 microns, several fold higher than when using "naïve" motors.

To summarize our results for various active transport designs, a comparative chart for the maximal observed lengths is given in Fig. 3.5.

3.4   Discussion

When considering biological active transport realized by molecular motors instead of passive diffusion, one typically envisions a conveyor-like delivery of the materials to the "construction site". Without active transport, as diffusional flux and the resulting elongation rate decrease with the length, retrograde flow plus depolymerization equalizes polymerization at the stationary length, and filopodial elongation stops [15]. With active transport, there is the possibility that forward flux exceeds that of retrograde flow, in which case there is no steady-state, and conveyor-like delivery with linear growth would last indefinitely.

Even when the motor-based actin forward flux is barely below the retrograde actin flux, it will still dominate the dynamics, which will exhibit a phase of linear growth for quite a long time (see Fig. 3.4). Diffusional flux will be gradually decreasing, until the total flux becomes equal to retrograde flux, which will happen when the filopodium has already grown long, as in the presented case with Ena/VASP.

If one were to write a diffusion equation for G-actin concentration [62], an addition of motor-based flux would be represented as a convectional term. From the arguments above,

the relative magnitude of this term with respect to retrograde flow is crucial, as it sets the asymptotic growth regime.

Our findings indicate that for a conveyor-like delivery achieved by an active transport term of large enough magnitude, the process of transport has to be organized quite intricately. Even for a modest increase in stationary length, certain "rules" have to be observed. These rules are quite general, so one can think of them as the rules for effective active transport in polymerization-based protrusions. To follow the rules, 1) motors should be kept from sequestering freely diffusing cargo, and 2) "rails" of the carriers should be kept clear of "clogging" by empty motors.

In our simulations, the directional flux due to active transport was almost always less than backward flux due to retrograde flow, therefore, diffusion still set the stationary length: that is when the diffusional flux thins out such that in sum with the motor-based flux it balances the retrograde flow. Surprisingly, it was only a narrow area of parameter space that yielded the "conveyor" picture of delivery, even in the artificial control simulations with sequestration and clogging being turned off. Perhaps this regime of transport would have been enhanced by simulating transport stoichiometry of 8 G-actins per motor instead of 2, leading to flux amplification. Nevertheless, we did not observe a "perpetual conveyor" in our speculative, yet plausible scheme with Ena/VASP. However, even achieving a transient, yet prolonged conveyor-like behavior required intricacy in that Ena/VASP needs to have both bundling and scaffolding (adaptor) functions. Interestingly, Ena/VASP is known to bundle filaments, could possibly act as a scaffold via its G-actin binding domains, and has been observed as myosin X cargo.

Thus, the balance between active transport-based flux and backward flux due to retrograde flow can dramatically affect filopodial growth. In this paper we considered a hypothetical active transport of G-actin to promote elongation. To the same purpose, retrograde flow might serve as another convenient way for a cell to regulate the filopodial length, switching between different growth regimes. Reducing the retrograde flow several-fold will increase the filopodial

stationary length [15]. In some cases, the reported values of bulk G-actin concentrations are much higher, on the order of 100 μM [127], however, *in vitro* 100 μM pure G-actin polymerizes, leaving only $0.1 - 1$ μM monomeric G-actin [128]. To maintain a pool of unpolymerized actin *in vivo*, most of monomeric actin is typically sequesterd by special proteins such as thymosin-β4 [129]. Prior experimental and computational analysis found that the concentration for non-sequestered G-actin is on the order of 10-50 μM [130]. Even at the upper limits of polymerizable G-actin concentrations, this will only lead to several-fold larger filopodial stationary lengths [15]. However, in combination with another several-fold increase that we observed when using the postulated Ena/VASP based scheme of G-actin transport, this can take filopodia from sub-micron lengths to the lengths on the order of ten microns. The same factors also increase growth speeds. Decreasing retrograde flow rate would increase the growth speed, by up to 4 $\mu m/min$. The combination of the latter effect with enhanced G-actin flux due to active transport and high G-actin concentration, may be enough to result in high growth speeds of $\sim 10 \mu m/min$ observed in some experiments [60]. Thus, the cells that grow extremely long filopodia perhaps use multiple facilities to achieve that, including the down-regulation of retrograde flow, up-regulation of actin concentration and possibly active transport, explored in this work.

It would be interesting to test experimentally if G-actin is transported actively with myosin X and/or Ena/VASP utilized as an adaptor. For example, one might mutate the G-actin binding domain of Ena/VASP and monitor filopodial length. However, Ena/VASP needs to bind G-actin for its anticapping activity, so mutation may diminish filopodial length through disabling anticapping, thus obscuring any conclusions regarding transport of G-actin. Another possibility is to fluorescently label G-actin: to avoid an intense background of glowing F-actin, one could consider labeling a small fraction of actin, or more attractively, labeling DNaseI which binds to G-actin but not F-actin [131]. Concurrent labeling of myosin X or Ena/VASP with a second fluorophore might reveal colocalization with G-actin along the length of the filopodia, or per-

haps even FRET-based spectral changes, offering support for the active transport of G-actin. Such a colocalization study might rely on techniques used in the recent work demonstrating that espin1 and myosin IIIa are cotransported along the length of a stereocilium [132].

Compared to filopodia, the requirement for active transport in other parallel actin based structures may not be acute. For instance, stereocilia are maintained at very definite lengths, where this fine-tuning is indicative of regulation by a possible signaling subnetwork.The elongation rate in stereocilia is lower by two orders of magnitude than in filopodia [107]. While this lower actin flux can probably be sustained by diffusion, nevertheless motors Myosin IIIa and Myosin XVa are found in stereocilia [132–134]. Myosin IIIa has been observed to carry epsin 1 protein that has an anticapping role [132], which is similar to M10 carrying Ena/VASP. Interestingly, myosin IIIa provides a significant boost to elongation, and this boost is dependent on the WH2 domain of espin 1, which binds to actin monomer [135]. Therefore, it is possible that the boost in stereocilia length might be mediated by active transport of G-actin. Mutation in M15a, also found in stereocilia, results in deafness in mice and humans because of stunted length [134]. Consequently, M15a may also be implicated in stereocilia length regulation, although not necessarily as a monomer transporter. The lengths of microvilli are shorter than of filopodia, so active transport of G-actin is less likely. However, motors are still found in microvilli too: Myosin I motor attaches the actin filaments to the membrane to power the sliding against it [136]. The role of Myosin VIIa in microvilli is less certain, though in other organelles it transports various cargo including proteins, melanosomes, and phagosomes [137]. Our model is general in its treatment of motors, hence, analogous active transport schemes could be constructed for microvilli and stereocilia, although one needs to take into account the mechanical and structural differences among these organelles.

3.5 Methods

Here we report a stochastic treatment of filopodial dynamics based on the Gillespie algorithm [83] that builds on our previously developed model [15] to study motor-mediated active transport of G-actin monomers to the polymerizing end. Individual reactions in the Gillespie scheme included 1) diffusion of G-actin along the length of the filopodium (rate k_D), 2) polymerization and depolymerization at the barbed end as affected by stochastic membrane force (rates k^+ and k^-), 3) motor loading and unloading of G- actin (rates k_l and k_u) , 4) motor binding and unbinding to and from the filament (rates k_{fb} and k_{fu}) and 5) motor steps along F-actin filaments (rates $k_\rightarrow$ and $k_\leftarrow$). In the simulations with Ena/VASP acting as an adaptor, the motor could be only loaded with an adaptor. The adaptor could be empty, or scaffolding one or two G-actin monomers. In yet another set of simulations, there was additional reaction, irreversible sinking of the adaptor with simultaneous release of G-actin (if it carries any) to the cytosol.

3.5.1 Actin bundle

Filopodia are formed by an F-actin bundle that extends the cell membrane, forming finger-like protrusions along the leading edge of the cell [2, 98, 138]. Typical extension lengths in real filopodia are several microns with growth and retraction rates of 0.1 - 0.2 μm/s [85]. Emergence of a filopodium from the 3D actin mesh of the cell or the lamellipodium is not simulated here. Instead we start with a pre-formed filopodium of 80 to 600 nm length, the actual value having no effect on the steady state. Since there are typically $10 - 30$ actin filaments per filopodium, we used 16 as the fixed number of filaments [62]. Accordingly, filopodia tend to be $100 - 300$ nm in width [63], so we used a width of 150 nm and assumed rapid mixing in the transverse direction. Longer filopodia may be thicker to maintain structural rigidity, though the steady-state effect of increased diffusing species would be canceled by the increased filament number. Having a persistence length of about 10 μm, or longer if tightly cross-linked, the bundle of F-

actin in our model was assumed straight for simplicity. Buckling and bending may be modeled in the future.

3.5.2   Diffusion

Stochastic implementation of duffusion is adopted from our previous model [15, 105]. In essence, the space is split into a 1D sequence of compartments along the filopodium and diffusion is realized as reaction of stochastic hopping between the adjacent compartments. The rate of diffusion happens to limit filopodial growth [15], which prompted us to investigate the effects of active transport of G-actin along the F-actin filaments by molecular motors.

Given the linear dimension, molecules diffuse quickly with respect to molecule reaction times up to a certain length, called the Kuramoto length [82], which is on the order of 100 nm with physiological molecule concentrations. This length can be thought of as a mean "free" (without reacting) path of a protein molecule. For stochastic treatment of diffusion, we divided the filopodium into compartments 50 nm in height, on the order of Kuramoto length, to allow molecules to randomly hop in 1-dimension from one compartment to another at rates that correspond to typical diffusion rates ($D = 5\mu m^2 s^{-1}$ diffusion coefficient, or 2000 s^{-1} hopping rate), though diffusion rates within filopodia have not been measured. Varying compartment height has been shown to have little effect on simulation results [15]. Diffusion constant for all the diffusing species in our simulation was $D = 5\mu m^2 s^{-1}$.

A boundary condition at the filopodial base maintains a G-actin concentration of 10 μM, and consumption of G-actin from barbed-end polymerization establishes a base-to-tip gradient. G-actin concentration at the tip can be very low [15], which is one the motivations for stochastic simulation of filopodial dynamics. The rate of diffusion thus limits filopodial growth, which prompted us to investigate the effects of active transport of G-actin along the F-actin filaments by molecular motors.

3.5.3 Retrograde flow

Since the depolymerization rate at the pointed end is slow, the limitation of filopodial length is mainly controlled by retrograde flow, where the entire actin bundle moves backward at a constant velocity [2,62]. The exact mechanism of retrograde flow remains speculative [12], but rates are likely subject to regulatory proteins. Here we did not consider variations in retrograde flow rates but instead used 70 nm/s in all simulations, while experimentally measured values are $\sim 10 - 200$ nm/s [16, 98, 100, 103]. Technically, we drag all the filaments with everything bound to them towards the filopodial base with a constant velocity of $v_{\text{retr}} = 70$ nm/s.

However, it is worthwhile to test this constant speed approximation. Retrograde flow is generated by several processes: treadmilling, membrane pushing downward and myosin based motions of the actin mesh inside the cell body that pull the filopodium backward. Although, the latter factor is likely the most important, it would be more realistic to take the contribution of the other factors into account as well. If an actin monomer is inserted between the filament tip and the membrane, it pushes both the filament and the membrane, and the filament flows back. Thus, the filopodium is only elongated by a fraction of monomer size, and the rest of it contributes to the treadmilling. Hence, the higher the effective polymerization rates are, the higher is the retrograde flow speed.

To include this coupling into out model, we additionally drag all the filaments back by a distance $r\delta/N$ (monomer size δ divided by number of filaments N multiplied by a coupling constant, r) after each polymerization reaction. The constant part of the retrograde flow speed still remains (v_r^0), although it is correspondingly smaller than the total speed in simulations where v_{retr} is constant. Fig. 3.7 shows the growth curves for $v_r^0 = 35nm/s$, $r = 0.3$, $k_u = 3s^{-1}$, $k_{fu} = 300s^{-1}$, $[M] = 1\mu M$ in the three cases: normal; no sequestration; no sequestration nor clogging. These simulations show that the qualitative difference between the dynamical regimes of the growth and changes in stationary lengths does remain the same as in the simulations with constant retrograde flow of $70nm/s$ and no coupling of retrograde flow to

polymerization.

3.5.4   Motors

To investigate active transport of G-actin, we incorporated directed motors into our model.
They can "load" and "unload" a G-actin monomer and step along the actin filaments. To model
such a step in the Gillespie scheme, a motor on a filament reacts with an actin monomer 32.4
nm (this is the size of a motor step) away from its starting location along the filament at rates
of $k_{\rightarrow} = 50s^{-1}$ in the forward direction and $k_{\leftarrow} = 5s^{-1}$ in the reverse direction. Not knowing
the real biological values and effects of unloading rates (k_u), we explored rates ranging from
1 to 3000 s^{-1}. Similarly, we explored motor-filament detachment rates (k_{fu} from 0.1 to 300
s^{-1}). Loading and attachment were assumed to be diffusion-limited (10 $\mu M^{-1}s^{-1}$).

Several different scenarios may arise for motors during the simulation: Motors near the
tip (closer to the barbed end than 32.4 nm) can no longer step forward, but can (un)load actin
(with rates k_u and k_l), detach from the filament (with rate k_{fu}) or step backward (with rate $k_{\leftarrow}$).
Motors that are not attached to any filament also diffuse back and forth within the filopodial
cytoplasm with $D = 5\mu m^2 s^{-1}$. Those that are bound to a filament are dragged backwards
by retrograde flow, so they are in essence walking along a treadmilling "rope". Thus, the
retrograde flow speed is effectively subtracted from average motor walking speed (found from
reaction rates and the step size).

Under more realistic consideration, the treadmilling speed of a filament depends on the
elongation rate, and so does the retrograde flow. Most of the simulations reported here were
carried out with constant retrograde flow rates, thus, neglecting this coupling. However, as
elaborated in Section 3.5.3, when we explicitly introduce the coupling between polymerization
and retrograde flow processes, the same qualitative conclusions on the rules of active transport
are reached as without coupling.

3.5.5 Ena/VASP

Ena/VASP are transported by myosin X [17] and have G-actin binding sites, therefore they might serve as adaptors/scaffolds for the myosin-based G-actin transport. In our scheme, an Ena/VASP molecule can bind 1 or 2 G-actins (with the diffusion-limited rate, $10\,\mu M^{-1}s^{-1}$, one by one), release either one of bound G-actins (with the rate k_s, that we explored in 0.1 to 300 s^{-1} range), be loaded to a motor (diffusion-limited rate) or unloaded from a motor with the rate k_u.

The irreversible sinking of Ena/VASP at the tip, representing the filament cross-linking role of Ena/VASP, was also diffusion-limited.

3.5.6 Polymerization, Depolymerization

Actin filaments are polar in that polymerization occurs predominantly at the tip, the barbed end, while depolymerization occurs predominantly at the base, the pointed end. For actin bundles that are linked to fixed substrate or extracellular matrix, this "treadmilling" is responsible for motility by extension of filopodia, and cytoskeletal dynamics in general [87, 88]. Whether protrusion or retraction occurs depends on the equilibrium between polymerization and depolymerization rates. G-actin polymerization rates are increased when bound to ATP, depolymerization rates are increased when bound to ADP, and F-actin:ATP hydrolyses to F-actin:ADP, leading to aging [139]. In our model, we assumed that only G-actin:ATP is polymerized at the tip and only G-actin:ADP is depolymerized at the base, allowing us to account for effects of hydrolysis solely via rates of polymerization ($k^+ = 11.6\,\mu M^{-1}s^{-1}$) and depolymerization ($k^-$ = 1.4 s^{-1}).

Polymerization and depolymerization were treated stochastically in the Gillespie scheme, and the depolymerization rate was slow compared to polymerization. One (de)polymerization event changed the filament length by 2.7 nm, although the diameter of G-actin is 5.4 nm, since

F-actin consists of two protofilaments in a right-handed helix. Thus, we modeled F-actin as one protofilament with a polymerization step of 2.7 nm. The motor step of 32.4 nm, therefore, spanned 12 actin units.

Polymerization required a free G-actin monomer, so, before one could be incorporated into the filament, it had to have been unloaded from the motor or Ena/VASP.

3.5.7 Membrane force

While depolymerization is unaffected by membrane force, polymerization rate decreases with membrane tension [140]. The membrane height must be large enough to sterically accommodate a G-actin monomer at the barbed end, so membrane force and its effect on polymerization rate were derived from membrane height. A Gaussian distribution of membrane height with respect to filament tip position was used with a square root of variance of 20 nm [15]. At the average height, the typical force was 10 pN. The height with respect to each of the 16 filaments was recalculated after each reaction because membrane fluctuations are on the micro- to millisecond timescale, rapid compared to growth dynamics on the second scale [93]. Longer filaments experience stronger membrane force and polymerize more slowly than shorter filaments, on average. This negative feedback diminishes the heterogeneity in filament lengths [15]. It should be noted that a second mechanism of negative feedback is the gradient of G-actin concentration along the filopodial tube, providing higher availability of G-actin for shorter filaments [15].

3.5.8 Simulation scheme

In this work we have used the spatial extension of the Gillespie algorithm [15]. In the Gillespie algorithm, a simulation step requires two random numbers. The first determines the time step as influenced by the aggregate rate of all possible reactions in the scheme (diffusion of various proteins, (de)polymerization reactions, motor (un)loading of G-actin, and motor (de)attachment from F-actin). The second random number determines which of these reac-

tions will occur, as affected by the rates. After an event, the following are updated: time, species in each compartment, filament length (always affected by retrograde flow, affected by (de)polymerization if the event occurred), and membrane force on each filament. Thus, the Gillespie method allows for simulation and evolution of reactions over continuous time while accounting for effects of molecular noise [65–73, 80]. In summary, we modeled stochastically the effects of motor transport of G-actin on filopodial dynamics, with molecular level spatial resolution for motors walking on actin filaments.

Table 3.1: Model variables and parameters in the simulations with myosin motors and Ena/VASP.

| Mechanics	
Half actin monomer size	$\delta = 2.7nm$
Number of filaments	$N = 16$
Thermal energy	$k_{\mathrm{B}}T = 4.1pNnm$
Membrane force	$f = 10pN$
Diffusion rates (all species)	$k_D = 5\mu m^2 s^{-1}(2000s^{-1})$
Membrane fluctuation	$\sigma_d = 20nm$
Retrograde flow speed	$v_{\mathrm{retr}} = 70nm/s$
Chemical reaction rates	
Polymerization	$k^+ = 11.6\mu M^{-1}s^{-1}(21.8s^{-1})$
Depolymerization	$k^- = 1.4s^{-1}$
Motor loading	$k_l = 10\mu M^{-1}s^{-1}(18.8s^{-1})$
Motor unloading	$k_u = 1 - 3000s^{-1}$
Filament-binding	$k_{fb} = 10\mu M^{-1}s^{-1}(18.8s^{-1})$
Filament-unbinding	$k_{fu} = 0.1 - 300s^{-1}$
Step forward	$k_{\rightarrow} = 50s^{-1}$
Step back	$k_{\leftarrow} = 5s^{-1}$
Adapter taking G-actin	$k_s^+ = 10\mu M^{-1}s^{-1}(18.8s^{-1})$
Adapter releasing G-actin	$k_s^- = 0.1 - 300s^{-1}$
Bulk concentrations	
Actin	$C_A = 10\mu M$
Myosin X	$[M] = 0.1 - 300\mu M$
Ena/VASP	$[V] = 0.1 - 300\mu M$
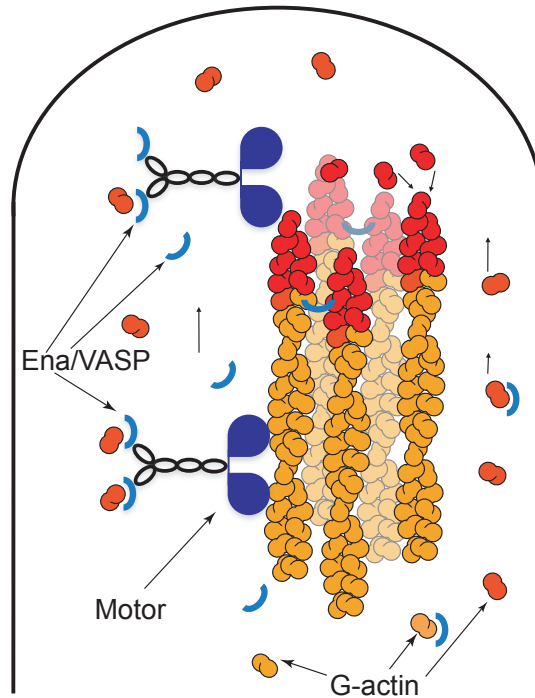
Figure 3.1: **A schematic representation of the filopodial tip in the model.**
A bundle of polymerizing actin filaments is enveloped by membrane which affects polymerization rates. Transported G-actin must dissociate prior to polymerization. Retrograde flow pulls filaments back with constant velocity. Myosin X motors travel the filaments in a directed fashion towards the barbed ends at the filopodial tip. Ena/VASP serves as a scaffold between G-actin monomers and motor molecules, and it is consumed near the tip due to cross-linking of the filaments.
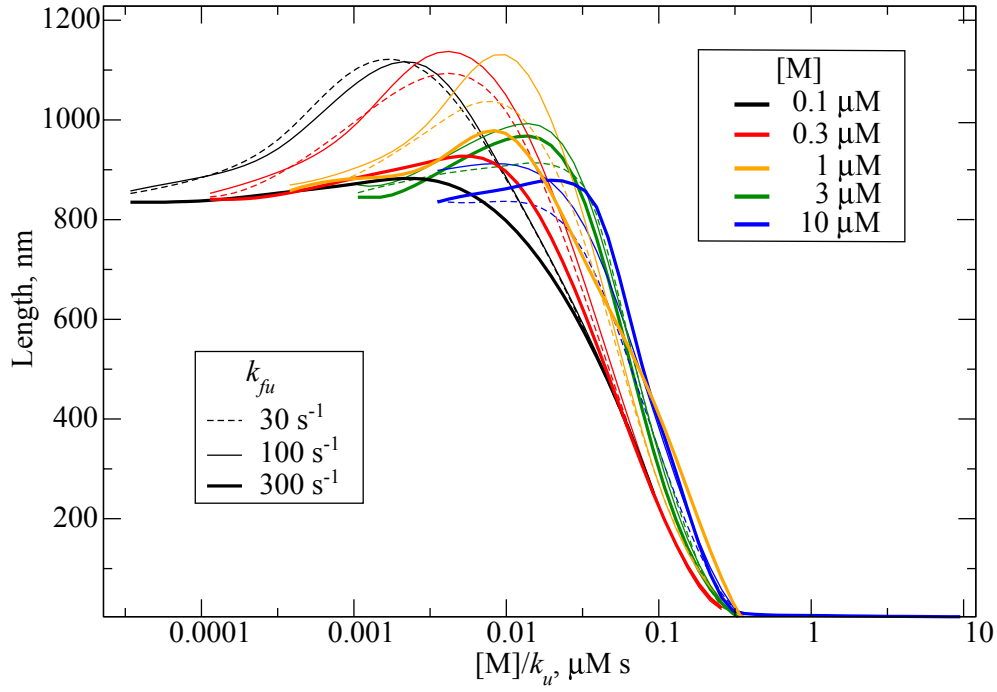
Figure 3.2: **Filopodial stationary length, as a function of model motor-related parameters.**
Bulk motor concentration is color-coded. Filament unbinding rate for the motors is equal
to 300 s^{-1} (thick solid lines), 100 s^{-1} (thin solid lines) and 30 s^{-1} (dashed lines). As the
motor unloading rate k_u essentially defines which fraction of motors carry actin, the length
dependence comes mostly from the ratio of motor concentration to unloading rate. For this
reason, the latter ratio is used as the variable on x-axis.

Figure 3.3: **Only in specific local regions of the parameter space do motors provide an increase in filopodial length.**
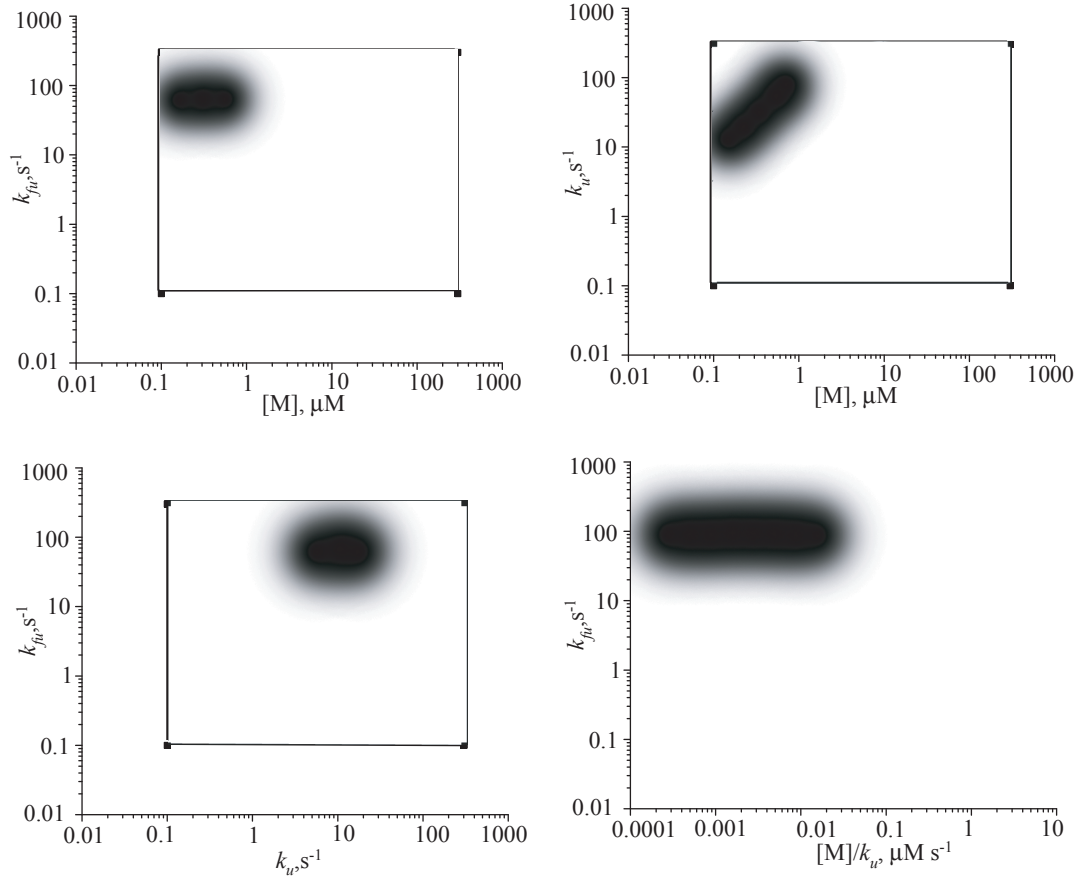These regions are always characterized by low motor concentrations. High motor concentration area does not lead to stable filopodia as most G-actin required for growth is sequestered.
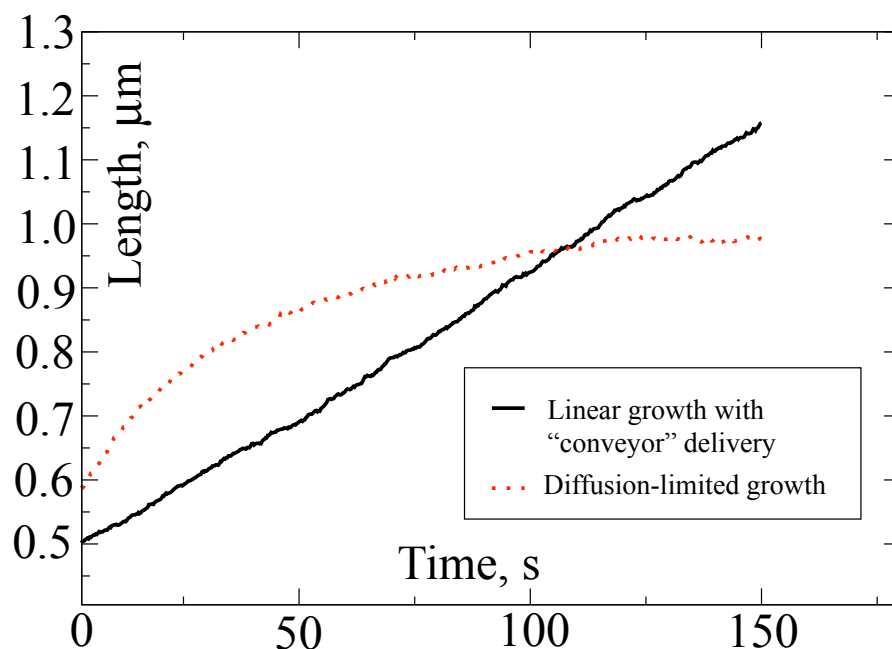
Figure 3.4: **Comparison between diffusion-limited and linear growth regimes**
These are two simulations with different parameters from the Ena/VASP set.
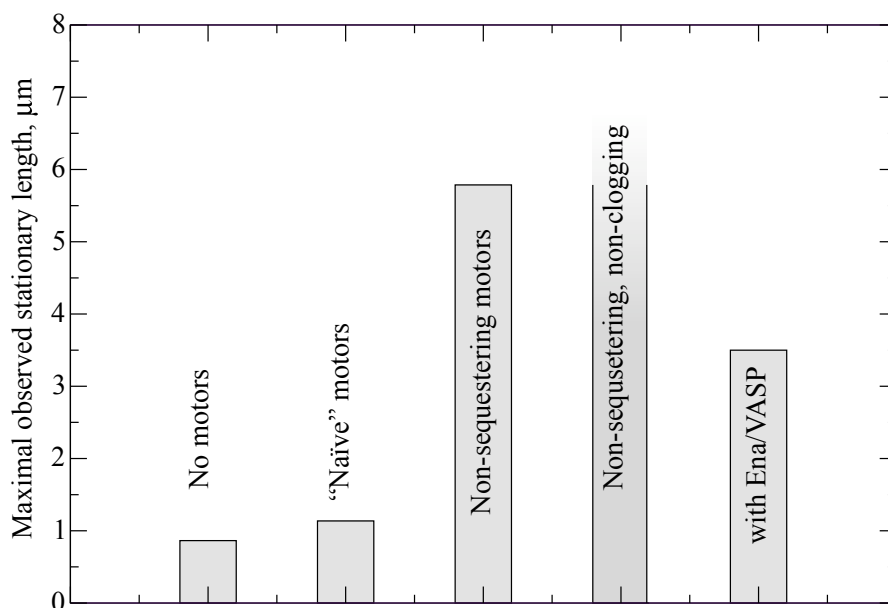
Figure 3.5: **The highest stationary lengths comparison.**
The largest observed stationary length in each set of simulations is shown. In case of simulations with artificial conditions, where motors do not sequester actin and do not clog the filaments, being sterically transparent to each other, a linear growth regime was observed, without reaching a stationary length.
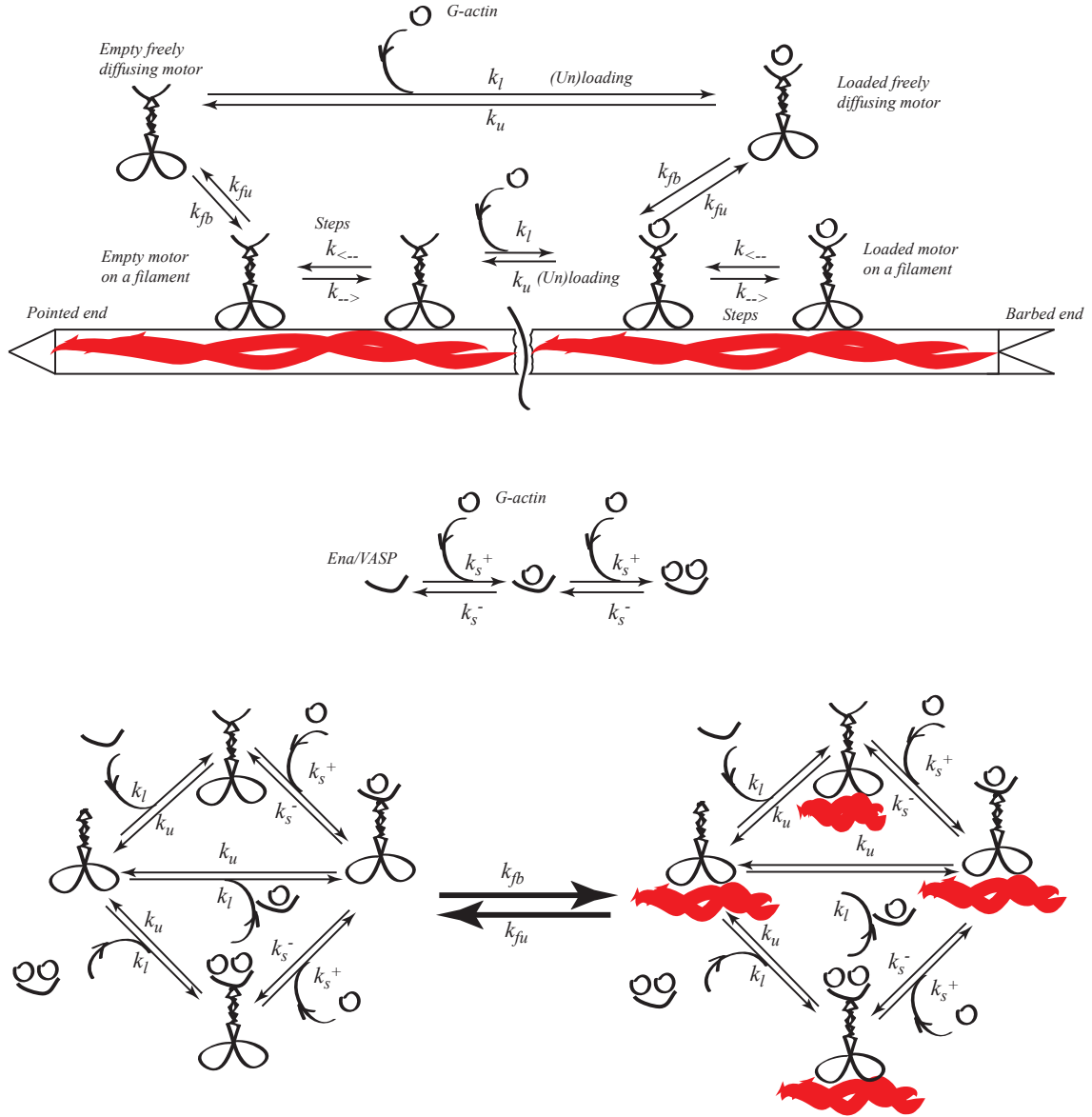
Figure 3.6: **A sketch of the motor transport processes implemented in the simulations** Normally, a motor could step only on a site with bare actin monomer unit. Thus, if it was closer than 1 step to the barbed end, the "step forward" reaction could not happen. Likewise it was not able to step on a monomer unit occupied by another motor. In the "no clogging" regime, these motors could change places. In the "no sequestration regime" reactions of loading and unloading in cytosol were not present (upper panel). Whether G-actin is transported directly by motors or by motors with Ena/VASP adaptors, it must dissociate from its transporter prior to polymerization

63
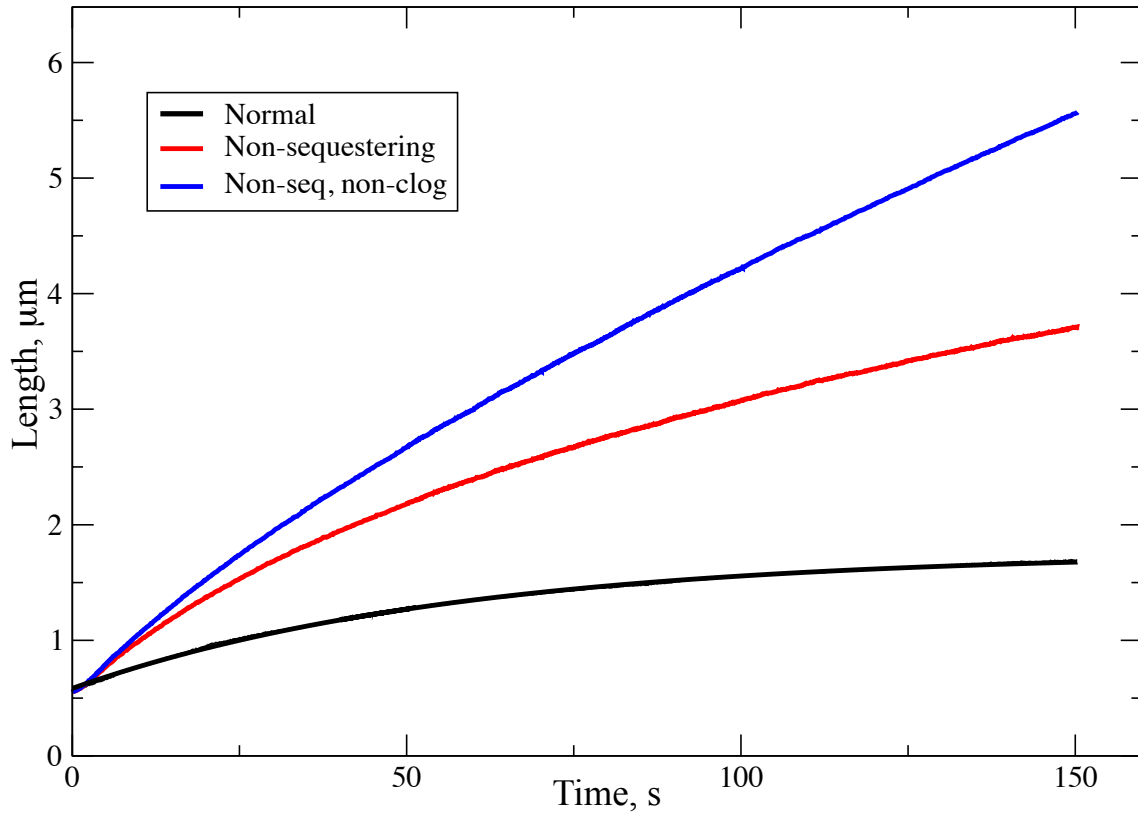
Figure 3.7: **The growth curves for a scheme where retrograde flow speed is dependent on polymerization.**
The parameters are $v_r^0 = 35nm/s$, $r = 0.3$, $k_u = 3s^{-1}$, $k_{fu} = 300s^{-1}$, $[M] = 1\mu M$. 3 curves correspond to normal (black); no sequestration (red); no sequestration nor clogging (blue). The influence of active transport bottlenecks remains the same as in the scheme with constant retrograde flow.

Chapter 4

High Resolution Approach to the Native State Ensemble Kinetics and Thermodynamics

Sangwook Wu, Pavel I. Zhuravlev, Garegin A. Papoian

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599

4.1 Abstract

Many biologically interesting functions such as allosteric switching or protein-ligand binding are determined by the kinetics and mechanisms of transitions between various conformational substates of the native basin of globular proteins. To advance our understanding of these processes, we constructed a two-dimensional (2D) Free Energy Surface (FES) of the native basin of a small globular protein, Trp-cage. The corresponding order parameters were defined using two native substructures of Trp-cage. These calculations were based on extensive explicit water all-atom molecular dynamics simulations. Using the obtained 2D FES we studied the transition kinetics between two Trp-cage conformations, finding that switching process shows a borderline behavior between diffusive and weakly-activated dynamics. The transition is well characterized kinetically as a biexponetial process. We also introduced a new one-dimensional reaction coordinate for the conformational transition, finding reasonable qualitative agreement with the two-dimensional kinetics results. We investigated the distribution of all the 38 native NMR structures on the obtained FES, analyzing interactions that stabilize specific low-energy conformations. Finally, we constructed a FES for the same system but with simple dielectric model of water instead of explicit water, finding that the results were surprisingly similar in a small region centered on the native conformations. The dissimilarities between the explicit and implicit model on the larger scale point to the important role of water in mediating interactions between amino acid residues.

4.2 Introduction

The native state of a typical globular protein is not a single static conformation but possesses rich intrinsic dynamics. In many cases these dynamics are essential for protein function. Examples include enzymatic catalysis [141], allosteric switching [141–143], protein-ligand binding (Mb-CO) [144], the change of antibody-ligand binding during maturation process [145], and

photorhodopsin photocycle at low temperature [146]. The main objective of this paper is to advance our understanding of the kinetics of conformational switching in the native basin of a globular protein. In particular, we address the question of whether native dynamics are activated, showing normal Arrhenius temperature dependence, or diffusive, similar to those of ordinary fluids. After reviewing the underlying energy landscape theory and reaction coordinates used in protein folding, we show a novel way to study the kinetics of protein conformational transitions by carrying out a Brownian Dynamics simulation on our two-dimensional free energy surface (FES) with two collective coordinates defined by the initial and final conformations.

A successful statistical mechanical approach to understanding protein folding is based on the concept of there being a funneled effective energy landscape [20, 21, 23, 48, 147–149]. The native state is thought to be a set of deep minima, in which the residual frustration between subtly different structural conformations leads to a rugged topography [20, 23, 48, 148, 149]. Residual entropy plays a similar role in spin glasses [20, 34, 48, 148, 149]. A protein at room temperature explores multiple conformations with different thermodynamically favorable contacts engendering this frustration [32, 34]. These native structural substates seem to be hierarchically organized into a tree with some ultrametric character[*], that, in principle, can be mapped onto the funnel (Fig. 1). Kinetics of transitions between the substates is also hierarchical with timescales dictated by the details of the distribution of energy barriers [20, 149–154].

By "energy" here we mean the energy of contacts. That, strictly speaking, is not an energy in the thermodynamic sense, because it includes the solvent coordinates in an implicit way. For example, hydrophobic interactions, though substantially entropic in origin, may be viewed as an "energetic" attraction between apolar residues. Even this effective energy of a protein conformation is still a function of hundreds or thousands of protein degrees of freedom [149].

---

[*]An ultrametric is a metric which satisfies the following strengthened version of the triangle inequality. Metric is a rule that defines a "distance" between any two elements of the set. If it is an ultrametric, then among three pairwise distances between any three elements two are equal and the third one is smaller. An ultrametric in a tree can be defined as the number of levels upward to the first common ancestral branch.

Often physical considerations permit phase space reduction through integration leaving only a few degrees of freedom that define then some very low-dimensional FES. Therefore, to avoid confusion (and also for historical reasons) the former representation of the landscape (funnel picture) is phrased in terms of energy, and the latter (low-dimensional FES) — in terms of free energy. The quantity and nature of the remaining coordinates determines the resolution of the FES. In particular, one may choose these so that the hierarchical organization of the substates and transitions will still be manifest. For instance, we may still see this hierarchy in non-trivial multiple timescale non-Arrhenius dynamics with an abundance of traps and barriers on the surface. Once the surface is constructed, studying the features of complicated protein dynamics consumes considerably less effort and permits a concentration on the kinetics of conformational switching. This technique can also be used to shed light on many interesting phenomena such as protein quakes [155].

In this work we constructed a FES, in two dimensions, for the native state of Trp-cage, a 20-residue protein that folds in $4\mu s$ [156, 157]. There are 38 NMR structures recorded for this protein in PDB code (1L2Y) [158]. Using the constructed surface we studied the transition between two NMR conformations, numbered 1 and 37, the most dissimilar of all 38 structures.

Among the plethora of possible order parameters in polymer physics and protein folding in particular, only those with a sufficiently high structural resolution can serve our purposes. One way to achieve such a high resolution is to define a coordinate with respect to a particular conformation A. A widely used coordinate is the root-mean-square distance (RMSD) between the corresponding atoms (usually, of the backbone, C_α). Another possible order parameter has evolved from spin-glass energy landscape theories and is a generalization of what is called the "overlap parameter" [159]. In case of polymer chains, the latter is a fraction of contacts that the two conformations have in common [42]. This parameter is often called Q, and may be defined as,

$$Q = \frac{1}{N} \sum_{ij} \exp\left[-\frac{(r_{ij} - r_{ij}^A)^2}{2\sigma^2} \right] \tag{4.1}$$

where r_{ij} are the distances between i-th and j-th atom in conformation of interest, r_{ij}^A is the same for the conformation A with respect to which the Q value is defined (usually, it is the native state), and normalization factor N is equal to number of pairs of atoms whose positions are defining the conformation [160]. The similarity index Q changes from 1 (for the conformation A itself) to 0 (when there is no resemblance to A). Indeed, the gaussian in Eqn. 4.1 suppresses the contribution of a pair of atoms if the distance between these atoms (r_{ij}) is very different from that in conformation A (r_{ij}^A). If one replaces the gaussian with a rectangular peak and, in addition, only includes pairs of proximal atoms ($r_{ij} \lesssim 5\text{Å}$), Q will turn into a fraction of shared contacts (fraction of native contacts when A is the native state).

Though RMSD might seem more familiar and natural, in some cases Q is preferable. If one imagines, for example, two conformations having two α-helices, that are close in one of them and apart in the other one, then RMSD between them will be very large, suggesting no structural similarity, while Q will still show the similarities of individual helices. Another nice feature of Q is that σ in its definition allows one to control the resolution of the order parameter, and one may tweak the notion of conformational similarity according to the particular questions under consideration. Yet another approach would be to use the fraction of shared dihedral angles of the backbone. Like Q, this parameter has shown considerable correlation with the strata of the folding funnel [19] when defined with respect to the native structure. Along with the native structure RMSD and fraction of native contacts or hydrogen bonds, the radius of gyration has also been used as a coordinate, since it is also correlated with folding. However, this measure lacks resolution inside the native basin which is the focus of our work, as further elaborated below.

Q is a good coordinate for protein folding because the funnel makes it possible to stipulate

that motions in directions transverse to Q will usually be fast, and therefore the motion along Q (folding) is nearly adiabatic [161]. A 1D profile is then quite appropriate. But if there were dynamical variables weakly correlated with Q but evolving on similar timescales, the information yielded by a 1D profile would be of very limited usefulness. In particular, studying the dynamics of allosteric switching may require higher dimensional order parameter space which allows the possibility to see multiple transition paths separately rather than integrated into an averaged one. A number of studies constructing 2D FESs have been reported in the context of protein folding. [153, 162, 163].

A natural choice of two dynamical variables having sufficient structural precision to provide a surface that allows the examination of the transition between two states 1 and 37 would be Q_1 and Q_{37}, that is, the similarities to states 1 and 37 respectively. Although Trp-cage is not an allosteric protein it nevertheless serves as a good test case for a general method that we develop here. In these coordinates the whole phase space is represented by a square $\{0 \leq Q_1 \leq 1; 0 \leq Q_{37} \leq 1\}$ (Fig. 2). The $Q_1 = Q_{37}$ diagonal of this square can be paralleled to a Q-axis in 1D free energy profile of protein folding. With our technique we pick two extremely close points on the 1D free energy profile, near the folded state ($Q = 1$), that correspond to models 1 and 37, and "pull them apart", unfurling the profile to a surface on a square and greatly increasing the resolution in the vicinity of reference states 1 and 37. On the (Q, R_g)-plot these two points would nearly coalesce, but we zoom in the region between and around them, leaving the unfolded state unresolved. All the unfolded structures are found close to the origin of the square.

We obtained the FES through extensive (over 1 μs) all-atom explicit water simulations in CHARMM force field [164] using 2D weighted histogram analysis method (WHAM) [165]. Subsequently, we ran damped Brownian dynamics starting from the basin of model 37, looking for the first passage time to the basin of model 1. Analysis of trajectories, first passage time distributions and temperature dependence of switching kinetics suggests a nearly free-

diffusional flow, with shallow traps weakly modulating transition dynamics. About half of the trajectories undergo partial unfolding down to $Q \approx 0.8$ and visit a number of local traps. We explored a one-dimensional order parameter for describing the kinetics of this transition. Brownian dynamics on this one-dimensional free energy profile yields FPT-distributions qualitatively similar to those obtained from the two-dimensional calculations.

We mapped the other 36 NMR structures onto the surface. This is an initial step in ranking them by free energies and organizing into an hierarchy. Having identified two deepest basins we compared structurally the corresponding conformations, highlighting stabilizing contacts and interactions. We also found that entropy and energy surfaces are much more rugged with steeper and higher rises and falls than that of free energy, which we attributed to fast solvent degrees of freedom. Lastly, we repeated the FES calculation with a simple dielectric solvent model (DSM) obtaining surprisingly similar results for the highly native part of the surface. Comparing the differences, however, showed the importance of water in stabilizing the folded state.

4.3 Methods

The Q values among 38 NMR structures were calculated using Biochemical Algorithm Library (BALL). [166]. All MD simulations were carried out using the LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) package [167] with CHARMM27 protein-lipid force field [164]. In DSM, the system with dielectric constant 80 was heated up to 282K and equilibrated for 800 ps using targeted MD to maintain the NMR structures. In the explicit solvent model, the protein was solvated with 2,275 TIP3P water molecules and the counter ions, $5Na^+$ and $6Cl^-$, in a $50 \times 50 \times 50$ Å^3 water box. The SHAKE algorithm was used to restrain the hydrogen bonds. Minimization was performed in two steps. First we minimized energy of water with protein fixed for 10,000 steps with a conjugate gradient method. Then,

using NAMD suite [168], we minimized the energy of the whole system for additional 10,000 steps. Using the **charm2lammps** perl script, the initial input file was generated for LAMMP, and subsequently, a NVT simulation was carried out for 12 ps, followed by an NPT simulation for 60ps with targeted MD. In the final production phase, 1.2 ns long simulations were carried out for each of the 914 WHAM windows. In subsequent analysis, the first 200 ps were discarded. The 2D umbrella potential used was $V_{umb} = k_1(Q_1 - Q_1^0)^2 + k_{37}(Q_{37} - Q_{37}^0)^2$. The number of windows, spirng constants and simulation time in each window were chosen on the basis of good overlap between neighboring windows. In other words, for each window, the rate of going to the areas covered by neighboring windows multiplied by the simulation time for a window should be considerably greater than unity. It turned out that the spring constants staisfying this criterion were in the range of from 11.2 kcal/mol/$Å^2$ to 72.8 kcal/mol/$Å^2$.The cutoffs of Coulombic and Lennard-Jones interactions were 10.0 Å. The trajectories were recorded every 3 ps in each window.

Brownian dynamics was carried out using Heun's method for the difference scheme [169]. FES was interpolated in MATLAB. Gradients were calculated from four nearest points in 2D. 40000 trajectories were used to obtain each FPT-distribution.

4.4 Results and Discussion

Since the configurations of side chains play a key role inside a native basin (Fig. 4.7), we included side chain carbons atoms (78 atoms total that include C_α, C_β, C_γ, C_δ, C_ε, and C_z) in the definition Eqn. 4.1 of Q. Fig. 2 shows the computed 2D free energy surface (FES) for the native state of Trp-cage peptide in terms of two similarity indices, Q_1 and Q_{37}. These two conformations (1 and 37) were chosen on the basis of being the most structurally dissimilar among the 38 NMR structures: they belong to different classes (structural basins) according to the hierarchical clustering [170] and have the smallest Q between them. Model 1 corresponds

to ($Q_1 = 1.0, Q_{37} = 0.889$) and model 37 – to ($Q_1 = 0.889, Q_{37} = 1.0$). All the different models of Trp-cage reside within the rectangle of $0.88 < Q_1 < 1.0$ and $0.889 < Q_{37} < 1.0$. The deepest microbasin on the whole FES is that of model 1 followed by model 16 as the next deepest.

4.4.1 Stabilizing structural features

A representative structure from the microbasin of model 1 is shown in lower right corner of Fig. 2 (this structure is almost identical to NMR model 1). Comparing this structure to model 16, which is the second lowest in free energy, reveals structural features that stabilize model 1: i) The 3-10 helix is not present in model 1, which is quite consistent with previous result by Zhou [171]. ii) Arg-16 is shifted towards Trp-6. This hints at the presence of a typical cation-π interaction between Arg and Trp, which are the likeliest amino acid residues to be involved in such interaction [172]. The disappearance of 3-10 helix (residues 11-14) may result from the side chain movement of Arg-16, which is adjacent to the helix and likely contributes to its sta- bility. iii) Tyr-3 is rotated in model 1, which is favorable for intramolecular interaction between Tyr-3 and Trp-6. Thus, the main factors that lower model 1's free energy are the energetic con- tribution from the interaction between Trp-6 and Arg-16, the intramolecular interaction due to a ring rotation of Tyr-3 and the increase of entropy due to the loss of 3-10 helix structure. The backbone RMSD between model 1 and model 16 is approximately 0.8 Å.

4.4.2 Transitions between microbasins

Protein dynamics have been thoroughly analyzed in many prior works [173–175], often us- ing very diverse viewpoints on the nature of dynamical transitions. Some techniques, such as normal mode analysis, treat proteins as solids possessing vibrations and phonons [176–178]. Another viewpoint is to imagine activated hopping between different conformations separated by energy barriers [151, 179]. This can be paralleled to the dynamics of a supercooled liquid. Yet another possibility is that the motion is similar to flow of a normal liquid such that the

system spends most of the time in saddle points rather than in minima of the energy landscape. There have been indications that activated, or even glassy dynamics, are more relevant to proteins, at least at low temperatures [20, 21, 32]. In this work we investigate more deeply the nature of protein dynamics at room temperature. Knowing the dynamical regime is important when studying fluorescence intermittency [180] or allosteric regulation [141]. For this purpose we considered the transition between two reference conformations, 1 and 37. If the dynamics are activated they should be dominated by a few or one pathway between the reference states. Another issue we address in this section is the determination of the ruggedness of the surface and how local traps influence transition dynamics. The influence of traps should be considerable in the activated regime and small in the diffusional flow regime.

To describe conformational transitions, we ran multiple 2D Brownian dynamics simulations on the computed FES [161], starting from the basin of model 1 ($Q_1 = 0.97$ and $Q_{37} = 0.895$) (see section 4.3). Two trajectories are shown in Fig. 3a. Distribution of first passage times (FPT) for transitions from model 1 basin to that of model 37 is given in Fig. 3c (black line). Our analysis of trajectories suggested a lack of dominant pathway. Indeed, even the trajectories picked from around maximum of the FPT distribution (a typical trajectory is shown in white in Fig. 3a) linger in different basins, despite other similarities. The trajectories taken from the shoulder of the FPT distribution (an example is shown in magenta in Fig. 3a) differ completely. It turned out that almost a half of the trajectories fall into the shoulder of the FPT distribution (Fig. 3c). The observed multiplicity of transition pathways may be engendered either by trapping in different local traps (and abundance of traps on rugged surface causing the large width of the FPT distribution), or by the thermal noise if it is large enough to easily overcome the barriers (free diffusion). Fig. 3b shows a magnified part of the surface depicting a microbasin. Even though the trajectory might seem as freely moving, it noticeably lingers in this microbasin. We find that there is almost no trapping as such, although the influence of fine features of the surface is conspicuous. This observation, along with the temperature

dependence of mean first passage times discussed below, directly suggests that protein confor-
mational dynamics at room temperature shows a borderline behavior between nearly diffusive
(as in a normal liquid) and weakly activated (as in a weakly supercooled liquid).

To further support this conclusion, we performed analogous calculations on the same sur-
face, but at several different temperatures, comparing the corresponding FPT distributions. In
reality, the FES itself would change with temperature, and therefore, these calculations do not
yield direct information on the behavior of Trp-cage at those temperatures: their sole purpose
is to provide reference points for identifying the character of the dynamical regime. Higher
temperatures promote the role of freely diffusive motions, while lower temperatures promote
trapping. The FPT-distributions computed at 153 K, 282K, and 600K are given in Fig. 3c. The
FPT distribution variance grows, as temperature drops, but relative variance, or coefficient of
variation, decreases. The latter is a measure of distribution width; when $C_v < 1$ the distribu-
tion is considered "low-variance", when $C_v > 1$ the distribution is considered "high-variance".
Thus, the number of effective pathways drops with decreasing temperature, and at low temper-
atures a dominant pathway emerges ($C_v = 0.27$ at 51K). We observe at 282 K ($C_v = 1.09$) a
crossover between the low-variance and high-variance regime.

The Arrhenius plot for the mean first passage time ($\langle \tau^* \rangle$), which is shown in Fig.4, pro-
vides an alternative way to analyze protein conformational dynamics. At low temperatures
we observe normal Arrhenius behavior with $\langle \tau^* \rangle \sim \exp(E_A/kT)$. At high temperatures depen-
dence becomes characteristic of the diffusive regime (for diffusion in confined space), with
$\langle \tau^* \rangle \sim D^{-1} \sim T^{-1}$. Dashed and solid curves are the fits to these functions respectively. Room
temperature behavior is near the crossover between these two regimes, with activation energy
of about $0.7kT$, corresponding to weak local trapping.

This provides another view on the problem of single- and multiexponential behavior that
has been discussed in the context of both regular (with a barrier) and downhill protein fold-
ing [181, 182]. If $P(\tau^*)$ denotes a FPT-distribution, or probability density that the transition

time is τ^* then $F(\tau^*) = \int_0^{\tau^*} P(\tau)d\tau$ is the distribution function, or the probability that the transition has occurred within time τ^*. The survival curve, $1 - F(\tau^*)$, is then the probability that transition has not yet occurred after time τ^*. Though this curve can be fitted by a single exponential, especially at long timescales (where such behavior might be expected , for example, for diffusion in a confined phase space), it turns out that much better fit is achieved by a double exponential (Fig.4.8). This is reminiscent of the dynamical behaviors discussed in the context of barrierless, or downhill protein folding where the shorter timescale is called "speed limit" [181, 183].

Summarizing this section, the dynamics of switching between two conformations shows features of both "flow" (as in a normal liquid) and weakly "activated" dynamics (as in a weakly supercooled liquid). Also, the transition kinetics shows biexponential dynamical behavior.

4.4.3   Explicit Solvent model and Dielectric Solvent model

We repeated the 2D FES calculations for a system where explicit water molecules were removed, and, instead, $\varepsilon = 80$ was used for all electrostatic terms in the Hamiltonian. The FES from this simple dielectric solvent model (DSM) is given in Fig. 5. It is more symmetric along the diagonal than in case of explicit water. DSM surface has a basin in the region where both Qs are around 0.75, while that of explicit solvent model does not; this suggests that water molecules prevent the tendency for partial unfolding. This point is also bolstered by the presence of barriers of $\sim kT$ height scattered in the region of Qs between 0.8 and 0.9 on the explicit solvent surface and absent on the DSM surface. The presence of the basin in lower left corner of DSM 2D FES is also a result of the lack of hydrophobic effect, which leads to opening of hydrophobic cores composed of Trp-6, Pro-12, Pro-17, Pro-18 and Pro-19 [184]; the structures there (white box in Fig. 5) correspond to a loop structure L found in a work by Juraszek and Bolhuis [184] as one of the intermediates during Trp-cage folding. In the upper part, near the

basins of 1 and 37, surface features are well preserved even though dielectric solvent model is extremely simplistic. The difference between the explicit solvent FES and the DSM FES given in Fig.6 demonstrates this more clearly. As Qs decrease more differences appear between explicit and DSM FESs, highlighting the role of the hydrophobic effect, which is completely absent in the DSM simulations. The 2D FES differences are asymmetric with respect to the diagonal line, indicating that the hydrophobic effect favors structures on one part of the surface. To explore this suggestion, we picked several structures from different sides of the diagonal line, $(Q_1 = 0.90, Q_{37} = 0.81)$ and $(Q_1 = 0.77, Q_{37} = 0.76)$, and calculated the solvent accessible surface areas (SASA) for each residue (see Table 4.1). The SASA of the hydrophobic core for the first region turned out to be about 10 percent smaller which provides further support that the asymmetric shape of the FES is also hydrophobic in origin. It is an interesting question whether reasonable corrections for the differences between explicit and DSM FES could be provided by introducing a simple hydrophobic term into the DSM Hamiltonian. It will also be interesting to use Generalized Born model to treat electrostatics instead of DSM. Altogeter, it is quite possible that a free energy surface calculated using GBSA Hamiltonian [163] will be much more similar to that the of explicit solvent.

Finally, we consider physical motivation for the treatment of the hydrophobic effect as an "effective" energetic term in implicit solvent models. The entropic and energetic landscapes in our explicit water simulations are both very rugged, which we attribute to the solvent degrees of freedom (Fig.4.9). However, protein chain itself is not influenced much by this solvent energy ruggedness, because of the timescale separation between fast motions of the solvent and slower motions of the chain. Instead, the protein chain moves mostly adiabatically in the averaged solvent field, where the resulting free energy ruggedness is about ten times smaller and much smoother. Therefore, treating hydrophobic interactions as though they are energetic has reasonable physical grounds.

4.5   Conclusion

Two-dimensional FESs are a powerful tool in studying kinetics and thermodynamics of the native state ensemble. With proper choice of dynamical variables one may control the resolution in various regions of the phase space. For investigation of allosteric switching kinetics between two states, similarities to each of these states (generalized fraction of shared contacts) serve as useful collective order parameters. In this work we computed a two-dimensional free energy surface for Trp-cage. We mapped all 38 NMR structures onto the surface, ranking them in free energies and identifying most thermodynamically stable conformations. We pointed out the main interactions that enhance the stability of the deepest basins. Using the computed 2D FES, we studied the kinetics of transition between two Trp-cage native substates reported by a NMR study. We found that the native dynamics of Trp-cage is borderline between diffusional dynamics and weakly activated dynamics. At room temperature and higher, the corresponding survival curve is best fitted by a double exponential. Transition dynamics become clearly activated at lower temperatures, and diffusional at higher temperatures.

To explore the role of the solvent, we also computed 2D FES for the same system but with static dielectric medium instead of water. The explicit and DSM free energy surfaces were quite similar in the vicinity of reference conformations, where resolution of our technique is the highest. However, when protein conformations become somewhat less native, systematic differences between explicit and DSM models point to the role of the hydrophobic effect. Computing energy and entropy contributions separately showed the major role played by water and the necessity for calculation of the free energy as a whole, the latter being a result of cancelation of very large terms. Our method may facilitate rigorous construction of coarse-grained force fields based on free energies. It will be interesting to apply the technique introduced in this work to conformational transitions in larger proteins.
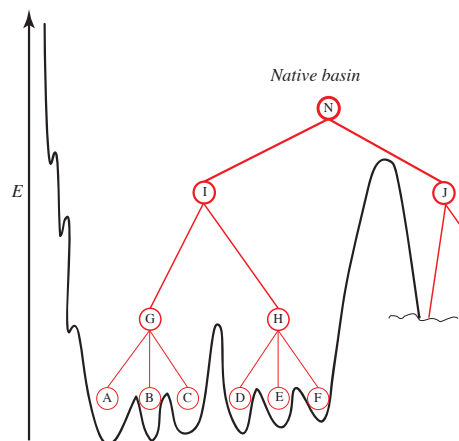
Figure 4.1: **A sketch of a hierarchical organization of native substates at the bottom of the energy funnel**

Native state on this schematic picture consists of two substates "I" and "J" interchanging on some timescale. On shorter timescale and higher resolution in energy each on of them also splits into substates. "I" splits into "G" and "H", that split further into yet another level of substates and corresponding microbasins.

(Q_1, Q_{37})	SASA of HP core	SASA of the whole protein
Model 1	955	1888
Hydrophobically favorable region		
(0.907, 0.807)	975	1943
(0.905, 0.809)	975	1952
(0.901, 0.802)	969	1931
(0.900, 0.808)	964	1901
Hydrophobically unfavorable region		
(0.774, 0.762)	1089	1936
(0.772, 0.763)	1120	1932
(0.773, 0.764)	1092	1920
(0.773, 0.761)	1038	2027

Table 4.1: Solvent Accesible Surface Areas ($\mathring{A}^2$) for structures indicated on Fig.4.6.
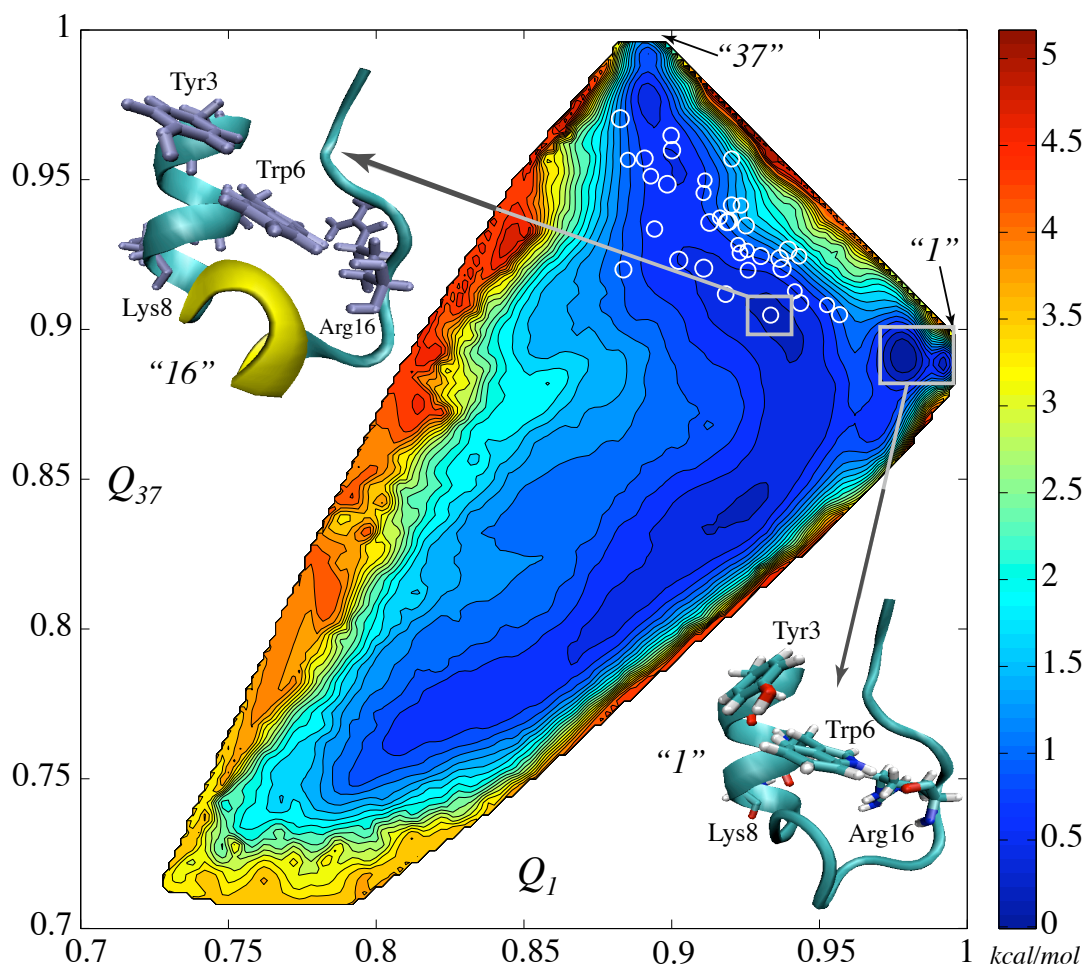
Figure 4.2: **Two-dimensional Free Energy Surface (FES) of Trp-cage native basin**
The free energy is plotted as a function of collective coordinates Q_1 and Q_{37}, indicating corresponding similarities to NMR models 1 and 37, was obtained using WHAM in 914 simulation windows (see section 4.3). An all-atom MD simulation of 1.2 ns in duration was carried out in each window. White circles, marking FES, represent the 38 NMR structures in the Q_1 and Q_{37} coordinates. Spacing between contour lines is 0.22 kcal/mol (0.37 kT). White rectangles mark the deepest and the second deepest basins corresponding to model 1 (lower right corner) and model 16 (upper left corner), respectively. The difference of $\sim 0.6\,kT$ in free energy is created by rotation of Tyr-3, cation-π interaction between Arg-16 and Trp-6 and disappearance of 3-10 α-helix]
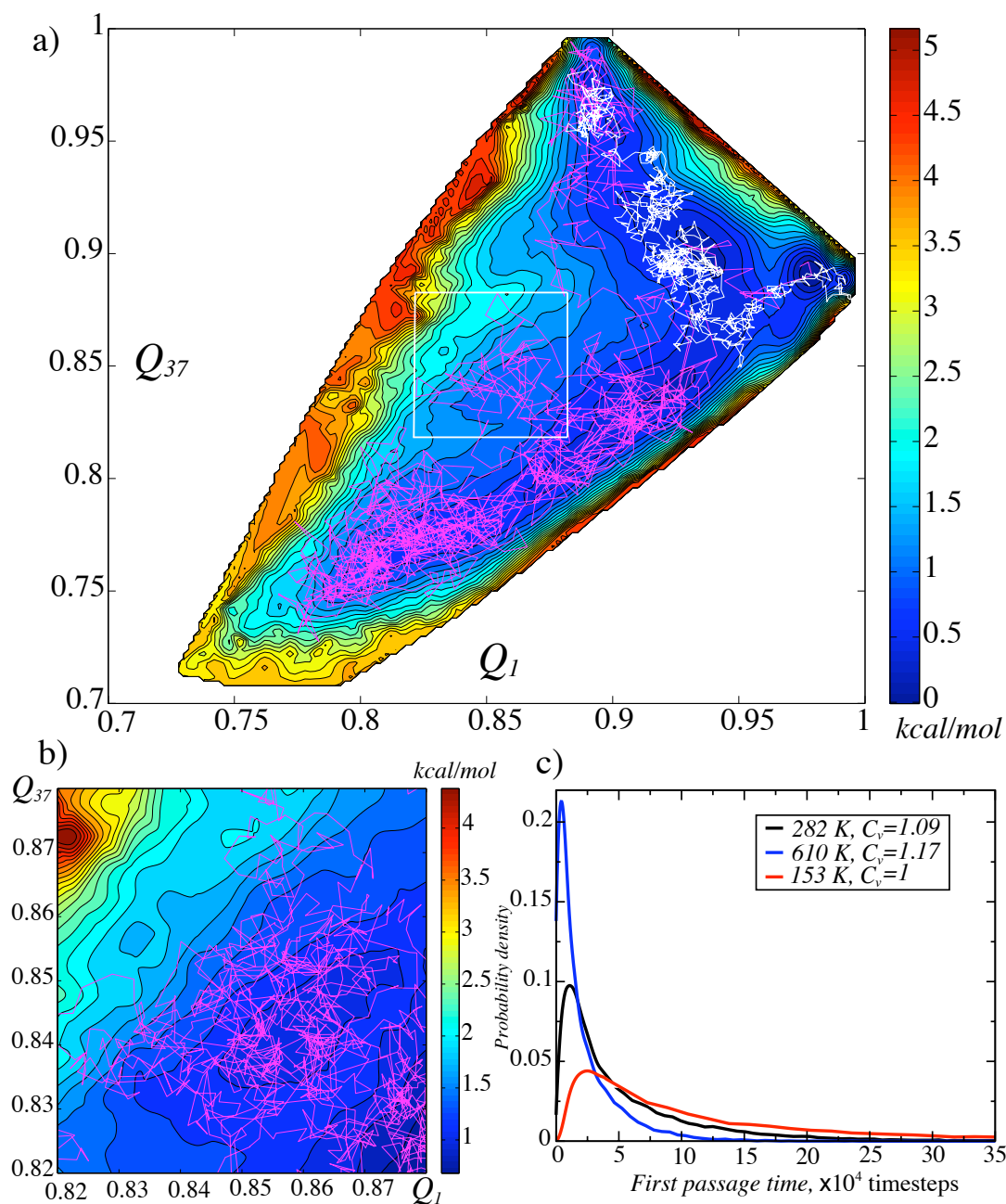
Figure 4.3: **Brownian trajectories on the free energy surface.**
Brownian dynamics of the transition between states corresponding to NMR structures numbered 1 and 37. a) Two sample trajectories: one (white) is from the peak of FPT distribution in (c), the other one (magenta) is from the shoulder; b) Part of the surface with trajectory shown in higher resolution revealing finer features of the surface and their influence on the trajectory; c) First passage time (FPT) distributions for conformational transition between 1 and 37. FPTs were computed at two additional temperatures to obtain the corresponding coefficients of variation, to provide reference for categorizing the nature of the dynamical regime at room temperature.
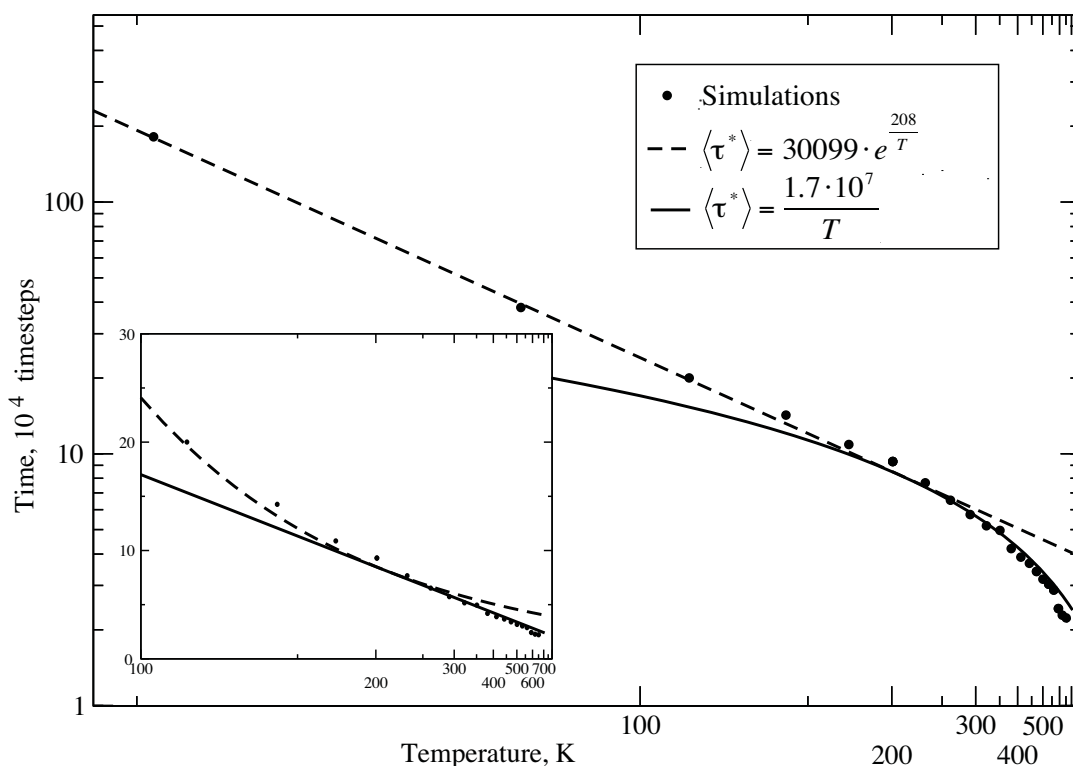
Figure 4.4: **Dynamical regimes of the transition revealed by temperature dependence.**
Arrhenius plot shows the dependence of the mean first passage time with temperature for the
conformational transition from "37" to "1" on fixed FES (computed at 282 K), where the FES
is temperature-independent. The inset zooms into higher temperature region using a semi-
reciprocal plot. Two regimes are clearly seen: exponential at low temperatures and linear at
high temperatures. The near room temperature (282K) result is near the crossover.
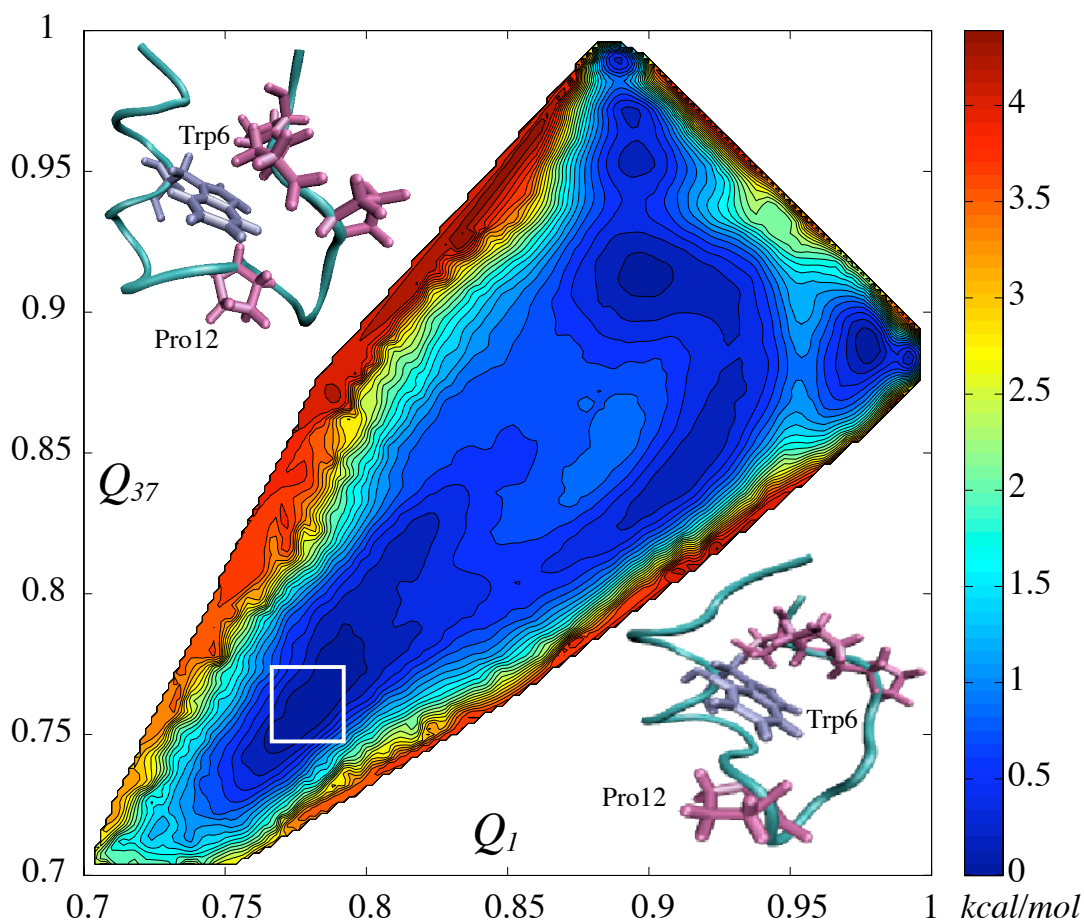
Figure 4.5: **FES was computed for the dielectric solvent model (see text)**

.

Two of the partially unfolded structures from the white square are shown in the corners. The lower basin is absent in the explicit solvent FES. On the other hand, a number of barriers have disappeared in the DSM FES. The vicinities of the reference points 1 and 37 are very similar between explicit and DSM FESs.
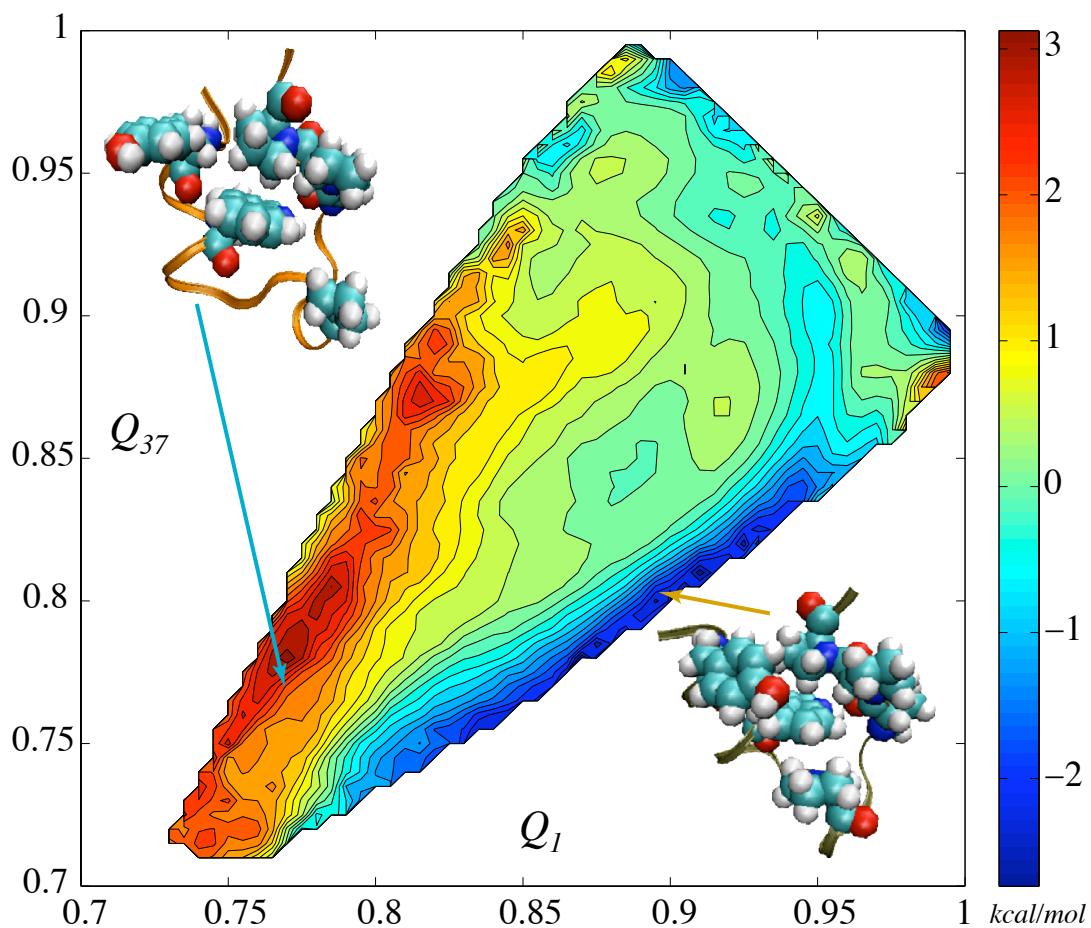
Figure 4.6: **The difference between explicit solvent FES and DSM FES.**
In the $Qs \lesssim 0.95$ region it mainly represents the hydrophobic effect. The structures in the corners are taken from the hydrophobically favorable and unfavorable regions. The hydrophobic core is shown in Van der Waals spheres, so that its opening is visually noticeable.
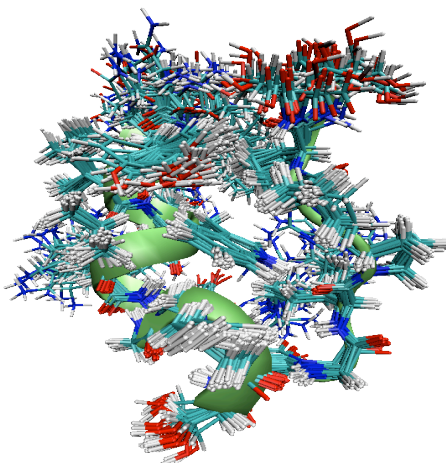
Figure 4.7: **38 NMR structures for Trp-cage superimposed.**
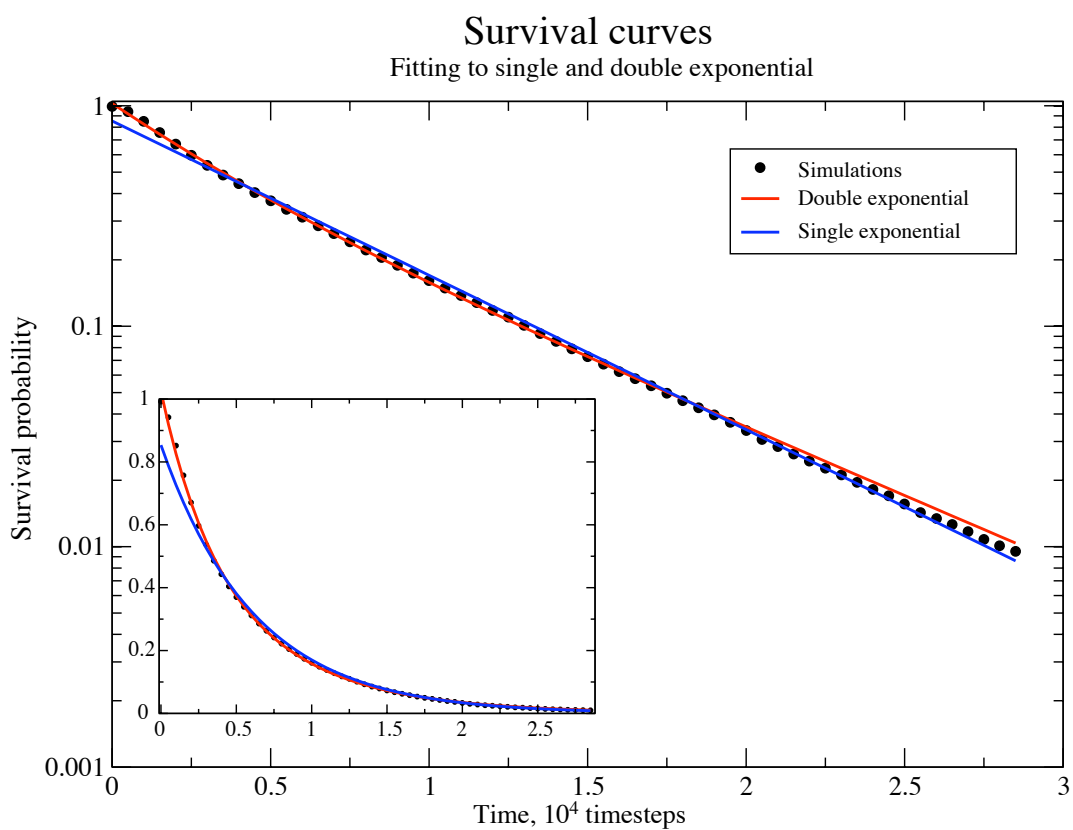
Figure 4.8: **The survival curve**
The probability that the "37" to "1" has not yet occured after time t fitted by a single (blue) and double (red) exponential. Inset shows the same plot on linear scale.
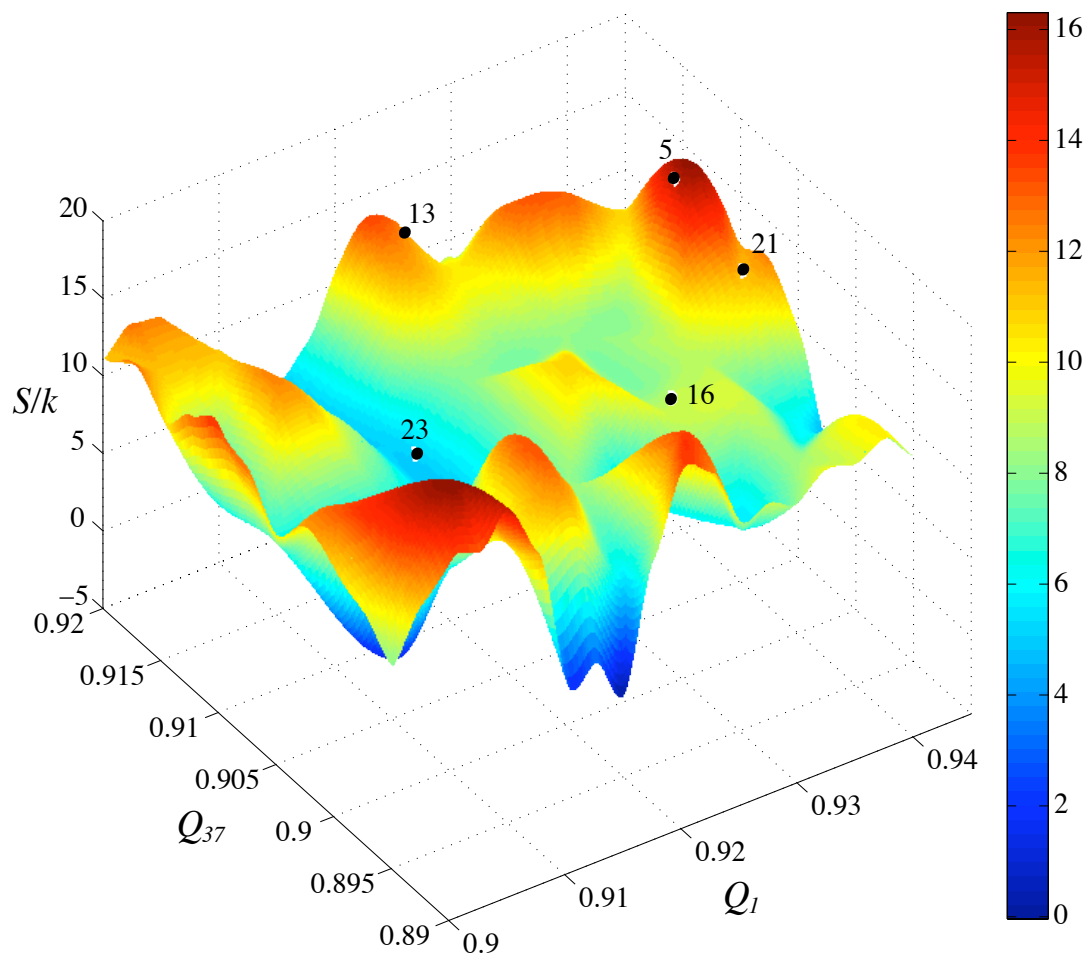
Figure 4.9: **Rugged entropy landscape around model 16 location spans 16 entropy units.**
The average energy landscape is very similar. Corresponding contributions to free energy is 10
times larger than change in free energy itself.

Chapter 5

Computing Free Energies of Protein Conformations from Explicit Solvent Simulations

Pavel I. Zhuravlev, Sangwook Wu, Davit A. Potoyan, Michael Rubinstein, Garegin A. Papoian

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599

5.1 Abstract

We report a fully general technique addressing a long standing challenge of calculating conformational free energy differences between various states of a polymer chain from simulations using explicit solvent force fields. The main feature of our method is a special mapping variable, a path coordinate, which continuously connects two conformations. The path variable has been designed to preserve locality in the phase space near the path endpoints. We address the problem of sampling the unfolded states by creating an artificial confinement "tube" in the phase space that prevents the molecule from unfolding without affecting the calculation of the desired free energy difference. We applied our technique to compute the free energy difference between two native-like conformations of the small protein Trp-cage using the CHARMM force field with explicit solvent. We verified this result by comparing it with an independent, significantly more expensive calculation. Overall, the present study suggests that the new method of computing free energy differences between polymer chain conformations is accurate and highly computationally efficient.

5.2 Introduction

Within the energy landscape paradigm, the protein native state is naturally viewed as a multitude of nested conformational basins, that are dynamically explored during protein function [20, 34, 185–189]. This functional landscape represents only a small fraction of the larger folding landscape – which includes denatured conformations [185] (see Fig.5.1). On the scale of the whole folding landscape, it is possible to describe folding dynamics through the statistical properties of the landscape. However, in the case of protein functional motions and native dynamics, the specific details of the functional landscape play an important role, necessitating detailed characterization of the landscape at a relatively high energetic resolution, corresponding to the structural resolution of $\sim 1\mathring{A}$. For example, such topographical maps [185] may

88

be needed, to investigate transitions in allosteric proteins, which undergo global conformational rearrangements upon local perturbation such as ligand binding. In some cases, allosteric switching is thought to modulate enzymatic rates [141]. Thus, elucidation of functional landscapes may help to understand how targeted point mutations influence catalytic activities [54] and may shed light on large scale phenomena, such as molecular motor functioning [190].

The energy landscape is a function of a large number of conformational and solvent degrees of freedom. In practical applications, the landscape is projected into one or several collective degrees of freedom, to allow physically meaningful interpretation of the chain dynamics. The present work provides the solution for a simpler problem: how to calculate the free energy difference between two specific conformations, *A* and *B*, of a polymer chain in a simulation with explicit solvent? Solving this problem is a step towards building a reduced representation of energy landscape and would not only help shed light on the biological processes mentioned above, but it would also aid in the development of atomistic and coarse-grained force fields, by allowing researchers to compare the free energy differences among the same conformations computed with different force fields and representations.

The attempts for addressing the challenge of calculating conformational free energies of molecules and macromolecules have a long history. A popular molecular mechanics/Poisson-Boltzmann/surface area (MM/PBSA) technique is based on generating a representative set of conformations with explicit solvent and then removing solvent and estimating free energy as a sum of several terms [191]. This technique is based on several uncontrolled approximations that may potentially limit its applicability [192], such as reliance on continuum electrostatics calculations to estimate part of polymer's solvation free energy, where these types of estimates can sometimes be quantitatively inaccurate [193]. A similar method, ES/IS, avoids using the Poisson-Boltzmann equation, and instead collects statistical averages computed from explicit simulations [194]. However, some of the terms in the free energy ansatz are still estimated by employing implicit continuum models of the solvent [194]. Some newer techniques like the de-

activated morphing method [195], which is based on using a series of unphysical intermediates states between conformations *A* and *B*, use fully explicit solvent. The deactivated morphing method has been used to calculate the free energy difference between folded and misfolded states of human Pin1 WW domain [196]. However, high (thousands of $k_B T$) free energy differences separating the unphysical states require extremely thorough sampling, pointing to a potentially very fast growth of computational cost with the system size. Structurally based umbrella sampling techniques have also been employed [145, 197], however, certain technical problems elaborated below significantly limit their domain of applicability.

To the best of our knowledge, the technique presented in this paper is devoid of any of the drawbacks mentioned above. It is fully general, takes the solvent into account explicitly, does not rely on any uncontrolled approximation (other than those intrinsic to any particular force field) and allows tuning of the conformational resolution. The method is accurate and computationally efficient. In certain cases, the full free energy profile for a transition may be obtained rather than just the free energy difference between two conformations.

Let point *A* in the phase space of a polymer chain be defined by precise coordinates of all the atoms of the chain. Point *A* has a finite entropy, and therefore statistical weight, due to solvent degrees of freedom. Point *A* only represents a point in polymer's conformational phase space, but it is expanded to a small region in the full phase space of the system. Furthermore, physically meaningful questions most often imply that conformation *A* includes not only the point *A* in conformational phase space, but also some finite size locale around point *A*. The latter is often called a conformational basin. The basin size depends on the question of interest and relevant physical considerations: for example, it could be defined by extent of atomic vibrations or by the experimental error in determining the structure of *A*. It may also be defined by the features of the particular local minimum on the energy landscape, such as its width and depth compared with thermal energy. The method that we report here does not provide explicit constraints on structural similarity within a basin, as these will vary between studies. Instead, it

provides means to calculate the free energy difference between the *A* and *B* basins, once those are defined based on other physical considerations. Thus, by "conformation A" we mean some well-defined neighborhood of point *A*, its conformational basin, and by "free energy of *A*" we mean the logarithm of statistical weight of this basin.

Many techniques for calculating free energy differences (such as umbrella sampling) require the free energy of the system to be computed as a function, $F(\xi)$, of a dynamical variable ξ, where common examples of ξ include density, magnetization, and radius of gyration. It is defined by a set of phase space variables and reflects the state of the system at any moment in time. Umbrella sampling is a way to sample low-populated regions of the free energy profile $F(\xi)$ by restricting trajectories to the narrow regions of the profile with parabolic potential $U = k(\xi - \xi_0)^2$. These regions are called umbrella windows, and the name comes from parabolic shape of the potential [198]. Coming up with an appropriate scalar variable ξ (that we will refer to as a path coordinate) for the problem of a transition between two polymer conformations is a non-trivial task. In this paper we present a method that solves this problem. The path coordinate has to keep most of the relevant information contained in the multitude of conformational and solvent degrees of freedom and simultaneously discriminate between *A* and *B*. In addition, both conformations *A* and *B* must correspond to finite segments of the ξ space, which means that conformations similar to *A* must have ξ close to $\xi(A)$ and vice versa, a region of $\xi \approx \xi(A)$ must only contain conformations similar to *A* (the statement has to be true for conformation *B* as well). In some cases, the topology of the landscape or allosteric motions themselves can provide a good, physically meaningful collective coordinate [199, 200]. The new path coordinate that we propose works for general case. It is local in the conformational phase space near points *A* and *B* and is highly resolved in discriminating them. We tested the method on two conformations from the native ensemble of a 20-residue protein Trp-cage [158] at temperature 282 K. The resulting free energy difference between them was found to be 0.43 kcal/mol (0.77 $k_{\mathrm{B}}T$). We also calculated the same free energy difference with an independent,

more computationally expensive technique resulting in 0.45 kcal/mol (0.81 k_BT) and confirming the accuracy of the method within 5%.

Any path coordinate is destined to have a range of values that contains all the unfolded states which are equally unrelated to either *A* or *B*. The phase space volume corresponding to this region is huge. Nevertheless, in general it must be sampled to obtain a free energy profile between *A* and *B*, unless energy landscape of the molecule has some intricate self-averaging property. Under specific circumstances, this may be the case in proteins, for example, when *A* and *B* correspond to an actual allosteric transition, but in general the sampling of this region might be problematic, particularly considering the computational cost of all-atom explicit solvent simulation techniques on currently available computational resources. We have conceived a solution of this problem by creating an artificial confinement in the phase space that prevents the molecule from unfolding without affecting the calculation of the desired free energy difference. Along with introducing the new path coordinate, this constitutes a new technique which is the main result of the current work.

Summarizing, in this article, we report a fully general and computationally efficient technique for finding conformational free energy differences between various states of a protein chain from all-atom explicit solvent molecular dynamics simulations. We applied our technique to compute the free energy difference between two conformations of Trp-cage (NLY-IQWLKDGGPSSGRPPPS) native ensemble, using the CHARMM force field with explicit solvent. We compared the results with those derived from an alternative, independent method, which is computationally much more expensive, revealing the remarkable efficiency and accuracy of the proposed technique.

5.3 The path coordinate and confining of the trajectories

To chart the native state in high resolution it is necessary to use some distance measure, $s(X,Y)$, between the points of conformational phase space, to quantify similarity between any two conformations X and Y. Examples of such measures include root-mean-square-deviation of corresponding atomic coordinates (RMSD), contact order, fraction of shared contacts (q), and fraction of shared dihedral angles. With such measure it is possible to map the whole conformational state on a single variable, i.e. to define the variable for an arbitrary conformation X. This variable is the similarity $s(X,N)$ between X and a preliminary chosen specific conformation N. This idea is used in protein folding with N being the native state [159]. The $s(X,N)$ is then the coordinate that describes folding. Since in our problem the two states may be very similar, as it happens in the native basin, we need a much higher resolution, and the one taking into account the conformational changes transverse to folding (i.e. transverse to $s(X,N)$) [185]. One way to increase the resolution is to use two variables instead of one, $s(X,A)$ and $s(X,B)$, the similarities to two specific conformations [201], mapping now the conformational phase space onto a 2D-plane (see Fig.5.2). Depending on particular choice of $s(X,Y)$ this variable may have different ranges. In many cases it changes from 0 (X and Y are totally different) to 1 (X is the same as Y). Fig.5.2 assumes such a case: both $s(X,A)$ and $s(X,B)$ can change from 0 to 1, mapping thus the whole conformational phase space onto a square. Area near the origin corresponds to conformations highly dissimilar to both A and B ($s(X,A) \approx s(X,B) \approx 0$). If A and B are both folded states belonging to the native ensemble (for instance, two allosteric states), the origin will contain all the unfolded states (since these states are dissimilar from the folded states). A and B then will be similar ($s(A,B) \approx 1$) and close to the upper right corner of the square. Then the diagonal of the square ($s(X,A) = s(X,B)$) will correspond to the folding coordinate line, and motions perpendicular to this diagonal will be transverse to folding. Fig.5.2b illustrates such a case: A and B are two structures from the native ensemble of a small protein Trp-cage. The structural resolution of the native region (upper right corner) is much

higher than that of unfolded region near the origin.

The most obvious path coordinate would be

$$\xi_d(X) = s(X,A) - s(X,B),\tag{5.1}$$

but it lacks the aforementioned property of locality near A and B which is compulsory. We want $\xi(X) = \xi(A) \pm \delta\xi$ to only include conformations similar to A, so that $s(X,A) = s(A,A) \pm \delta s$ with δs being small*. However, the difference based definition of path coordinate ξ_d (Eq.5.1) permits arbitrary large changes to both terms in the difference as long as the difference itself stays the same. In other words, the whole strip that is highlighted in white in Fig.5.2b will contribute to free energy of conformation A, including the unrelated unfolded conformations near the origin.

We propose a path coordinate that remains local around the conformations of interest:

$$\xi_g(X) = e^{-\frac{(s(X,A)-s(A,B))^2+(s(X,B)-1)^2}{2\sigma_g^2}} - e^{-\frac{(s(X,B)-s(A,B))^2+(s(X,A)-1)^2}{2\sigma_g^2}}.\tag{5.2}$$

If this coordinate is visualized as elevation above the 2D-plane defined by $s(X,A)$ and $s(X,B)$, it corresponds to a positive gaussian peak of width σ_g centered on conformation B (with coordinates on the 2D plane $s(A,B)$ and $s(B,B) = 1$) and a negative gaussian peak centered on conformation A (with coordinates on the 2D plane $s(A,A) = 1$ and $s(A,B)$) (Fig.5.2c,d). In this coordinate constant elevation strips form local regions of width σ_g near points A and B. Note that it is not necessary for both gaussians to have the same width σ_g.

We chose a small protein, Trp-cage, to test our method. Trp-cage is one of the smallest known proteins (20 residues) with a set of native structures reported by an earlier NMR study [158]. We chose the two most dissimilar structures in this set as points A and B. Despite the fact that we had allosteric states in mind while developing the method, these Trp-cage states are

---

*Or more formally: $\forall \delta s > 0, \exists \delta\xi > 0 : \forall X : |\xi(X) - \xi(A)| < \delta\xi \Rightarrow |s(X,A) - s(A,A)| < \delta s$

not expected to represent deep minima, but are simply used to test our approach. Furthermore, although allosteric states are typically minima with a barrier separating them, our method is more general and can be applied to any two arbitrarily defined conformational states, even if they are not minimum energy structures. Our technique allows computation of the correct ratio of thermal probabilities to find the system in either of these states, or the free energy difference. Free energy differences between conformations which are not deep minima may be used for example to gauge the accuracy of coarse-grained force-fields, by comparing with the corresponding results from atomistic simulations. Likewise, our technique may be used with many different similarity measures, $s(X,Y)$. In this work, we chose the fraction of common contacts $q(X,Y)$ to quantify similarity between structures X and Y, or more precisely

$$s(X,Y) = q(X,Y) = \frac{1}{N} \sum_{ij} \exp\left[-\frac{(r_{ij}^X - r_{ij}^Y)^2}{2\sigma^2} \right], \tag{5.3}$$

where r_{ij}^X and r_{ij}^Y indicate the distances between i-th and j-th atom in conformations X and Y respectively, and normalization factor N is equal to the number of atom pairs used to compare structures X and Y. In the example with Trp-cage we included carbons C_α, C_β, C_γ, C_δ, C_ε, and C_z (78 atoms total) in the summation. Gaussian function in Eq.5.3 smoothes the boundary between a "contact" and "no contact". Further in the text we use the following notation:

$$Q_A \equiv q(X,A), Q_B \equiv q(X,B), \xi_g \equiv \xi_g(Q_A, Q_B) \equiv \xi_g(X).$$

The comparison of locality between the previous, difference based path coordinate ξ_d and the newly proposed gaussian based path coordinate ξ_g is shown in Fig.5.3. Patches of different colors correspond to different windows of umbrella sampling (that keep the path coordinate localized). It can be seen from Fig.5.3a that the previously reported coordinate ξ_d indeed forces the trajectories to sample a stripe-like region of the 2D (Q_A, Q_B) plane. The ξ_d windows that contain conformations A and B also group with them unrelated, partially unfolded structures.

In comparison, when using the new path coordinate ξ_g, the window that contains A is local, as shown in Fig.5.3b, and the trajectory in this windows does not stray far from A, keeping the conformations unrelated to A from this window.

The new path coordinate ξ_g constitutes the essence of the technique reported here. However, there is another major feature that might be needed for efficient calculations under specific circumstances. Note that the conformational phase volume as a function of ξ_g is not constant: it is much larger in the region $\xi_g \approx 0$, which contains all the unfolded states and decreases rapidly towards the endpoints $\xi_g = 1$ and $\xi_g = -1$. In principle, the $\xi_g \approx 0$ umbrella windows have to be thoroughly sampled as well which may represent a problem with large proteins and explicit solvent force fields. However, if one is interested only in free energy difference between conformations A and B and not in the full free energy profile between them, this problem can be side-stepped. It is possible to confine the sampling trajectories inside an artificial "tube" that envelops a presumptive path between the states. This can be done by adding an appropriate confinement potential V_c to the Hamiltonian, $H' = H + V_c$. V_c should be chosen in such a way that the conformational basins of A and B are not affected ($V_c(A) \approx V_c(B) \approx 0$). Using the modified Hamiltonian, the free energy difference between the conformations A and B is

$$F'_A - F'_B = -\frac{1}{\beta} \ln \frac{\int\limits_{\Gamma_A} e^{-\beta H'} d\Gamma}{\int\limits_{\Gamma_B} e^{-\beta H'} d\Gamma} = -\frac{1}{\beta} \ln \frac{\int\limits_{\Gamma_A} e^{-\beta(H+V_c)} d\Gamma}{\int\limits_{\Gamma_B} e^{-\beta(H+V_c)} d\Gamma}, \qquad (5.4)$$

where $\beta = 1/kT$, Γ represents the whole conformational space and Γ_A and Γ_B indicate the phase volumes of conformations A and B. If we set $V_c = 0$ (below desirable marginal error, e.g. less than 0.01 kcal/mol) everywhere in Γ_A and Γ_B, then $F_A - F_B = F'_A - F'_B$, thus, V_c will not affect the free energy difference we are calculating. In our test example with Trp-cage we chose a V_c that won't allow the trajectories to unfold. On the (Q_A, Q_B) square this would mean preventing the trajectories from going towards the origin, keeping them in the upper right corner, corresponding to the native region. Thus V_c can be visualized as a wall of cylindrical

shape surrounding the upper right corner of the phase space square (Q_A, Q_B). Keeping the radius of the cylinder small would aid computational efficiency, but it should be large enough to not touch the conformational basins of *A* and *B* and to allow sufficient overlap between umbrella windows (see Methods Section and Fig.5.5).

5.4 Results

The free energy profile as a function of ξ_g is shown as a black solid curve in Fig.5.4b. This 1D profile was calculated using 108 umbrella windows that are each 1.2 ns long. High conformational entropy in the region $\xi_g \approx 0$ is the thermodynamic factor which tends to lower that region's free energy. However, we are mainly interested in obtaining the free energy difference between the conformations *A* and *B* (marked by rectangles). As discussed in introduction, these correspond to finite segments of the path coordinate ξ_g line. The size of these segments is not set by our method and must be chosen based on other considerations, such as magnitude of atomic vibrations. If the structures for *A* and *B* are obtained from experiment, the size may be defined by experimental precision. For a given segment size, though, our procedure provides a definite answer. It makes physical sense to choose a segment size that corresponds to a conformational basin, the local minimum on the free energy landscape, providing there is one. We chose the sizes to be $\Delta\xi_g = 0.2$. To calculate the free energy of a segment, we sum the partition functions of all the states within it: $Z_A = \int_{0.8}^{1} \exp(-\beta F(\xi_g))d\xi_g$ and $Z_B = \int_{-1}^{-0.8} \exp(-\beta F(\xi_g))d\xi_g$. The free energy difference $F_A - F_B = -kT \ln(Z_A/Z_B)$, turns out to be 0.43 kcal/mol ($0.77k_BT$).

To independently verify this result we also constructed a 2D free energy surface as a function of Q_A and Q_B (Fig.5.4a). 2D FES calculations are much more expensive computationally [165, 202, 203]. For these calculations, 922 umbrella windows are needed, 1.2 ns each resulting in over 1μs of total simulation time (compared to $\sim$ 100 ns simulation time for the 1D

profile). To compare free energies from 1D free energy profile $F(\xi_g)$ and from 2D free energy surface $F(Q_A, Q_B)$ we integrated the 2D FES numerically using

$$e^{-\beta F(\xi'_g)} = \int e^{-\beta F(Q_A, Q_B)} \delta(\xi'_g - \xi_g(Q_A, Q_B)) dQ_A dQ_B.$$

The profile obtained from the 2D surface $F(Q_A, Q_B)$ yields 0.45 kcal/mol ($0.81 k_B T$) free energy difference between the basins, to be compared with 0.43 kcal/mol ($0.77 k_B T$) obtained from 1D $F(\xi_g)$ calculations. The difference is within 5%, indicating that the method is highly accurate.

5.5 Discussion

The new method presented for computing free energy differences between polymer chain conformations has several advantages compared to previous approaches. The technique is general and does not involve calculations with unphysical states of the molecule (other than forcing the system to visit states that are poorly accessible thermally). It has adjustable structural resolution, that can be changed depending on the nature of the two conformations of interest. The resolution can be changed by increasing or decreasing the set of atoms whose positions enter into the definition of $s(X,Y)$. For instance, partially unfolded states of proteins would require coarser treatment, than the one in the example here: side-chain rearrangements should be considered as not changing the conformation, because they would occur on the same timescale as solvent motions.

The enveloping "tube" for the pathway between A and B is not as artificial construct as it may seem at first glance, at least regarding proteins. In real allosteric transitions proteins do not unfold (or do so only partially), which means a natural tendency to stay in the region of the energy landscape we are trying to sample. Thus, many proteins would naturally sample a well-defined path if the umbrella window sampling times are not too long, which would

allow escape over kinetic barriers surrounding the dominant path. In this case, an externally introduced confining tube would serve only as a "guard rail" for the trajectory rather than a wall that cuts a part of phase space off. In the example presented in the paper we used the following considerations to find a functional form for V_c that would approach such ideal case. First, the tube has to allow sufficiently many pathways connecting A and B, so that there is enough overlap between the umbrella windows. Then, the rate of transition between neighboring windows (with V_c on) multiplied by the simulation time inside a window must be larger than unity. However, this should not be an excessively large number, since it is a gauge for the sampling problem that the tube is meant to solve. Thus, some optimal width tube needs to be devised around the steepest descent path between the end points. Since this path is not known *a priori*, we probed the phase space with short time umbrella windows, thus "seeing" where the trajectory "prefers" to go (Fig.5.3b), until we observe a continuous path between A and B. The areas where the system spends most of its time are grouped around the steepest descent path, forming the shape of the tube (see Methods section). A more general procedure for constructing an adjustable confinement tube and making sure that its main purpose is to be a "guard rail" would allow the calculation of realistic transition pathways between the states, not just the free energy difference, with high computational efficiency. Thus, if the system naturally wants to go from A to B, the tube acts as a "guard rail", and we can recover the real transition pathway with a free energy profile along the pathway. In the opposite case, when system does not want to go from A to B, we may need to force it with the confinement tube, and then we recover just the free energy difference.

Root-mean-square-deviation of atom positions (RMSD) is widely used as a similarity measure between two conformations, s. Our method is formulated in terms of general $s(X, Y)$, so RMSD could also be used. The plot corresponding to Fig.5.2 would look somewhat differently (the whole quadrant instead of square with the native region near the origin) but with the same main features. We preferred to use fraction of native contacts $q(X, Y)$ for the following

reasons. q is a two-point parameter, comparing the distances between pairs of atoms, while RMSD compares the coordinates of each atom individually (after proper alignment). Use of q does not require this preliminary alignment of the structures. As the contact energy plays crucial role in proteins, q is more correlated with protein physics than RMSD. For example, if one imagines two conformations having two α-helices, that are close in one of them and apart in the other one, then RMSD between them will be very large, suggesting no structural similarity, while q will still show the similarities of individual helices.

In fact, umbrella sampling with the path coordinate ξ_d (Eq.5.1) and RMSD in the role of s has been used to calculate free energy difference between $A-$ and $B-$forms of DNA [197] and free energy mapping of allosteric switching between the open and the closed forms of adenylate kinase upon ligand binding [145]. As discussed above, this technique has the non-locality problem, that we have solved by introducing a new path coordinate in this paper. Along with the confinement idea, our technique is fully general, adjustable to high resolution, computationally efficient and does not operate on unphysical states of the system. The method can be used for an arbitrary pair of states, without the naturally existing transition path between them, like in the current work, to compute free energy differences between two protein conformations. In addition, a straightforward generalization should allow computing the whole free energy profile for the transition path between two states, when this transition occurs in nature.

5.6 Methods

All atomistic molecular dynamics (MD) simulations were carried out using LAMMPS [167] (Large-scale Atomic/Molecular Massively Parallel Simulator) using CHARMM27 protein-lipid force field with explicit solvent [164]. Trp-Cage was placed in a $50 \times 50 \times 50$ Å^3 box with 2,275 water molecules (TIP3P model) and the counterions, $5Na^+$ and $6Cl^-$ in order to mimic the physiological conditions. The system was prepared in NAMD using the standard

protocol [168]. The system was heated up to 282 K and equilibrated for 800 ps using targeted MD to keep the innate NMR structures. Next, NPT simulations were carried out in LAMMPS for 60 ps with targeted MD to bring the conformation to a specific umbrella window. Finally, for each of the windows, 1.2 ns long NPT simulations were carried out, where the last 1 ns was used for data analysis.

The confinement potential for the "tube" enveloping the trajectory is constructed as follows. In the case of Trp-Cage, the upper right corner of the (Q_A, Q_B) square is naturally devoid of states, thus, we only need to confine the trajectory from the side of lower values of Q_A and Q_B. We placed a cylindrical "wall" around upper right corner of conformation space square (Fig.5.4a), that confines the trajectories to the upper right corner. The "wall" (actually, a sharp step with finite width) was implemented by a hyperbolic tangent of a distance from the upper right corner of the phase space square $(Q_A = 1, Q_B = 1)$:

$$V_c = \varepsilon \left(1 + \tanh \left(k \left[(Q_A - 1)^2 + (Q_B - 1)^2 - \mu^2\right]\right)\right). \tag{5.5}$$

The parameters ε, k and μ, that were chosen as 10 kcal/mol, 5×10^3 and 0.135 respectively, satisfy the conditions of good overlap between umbrella windows, but, eliminate a huge number of intermediate states.
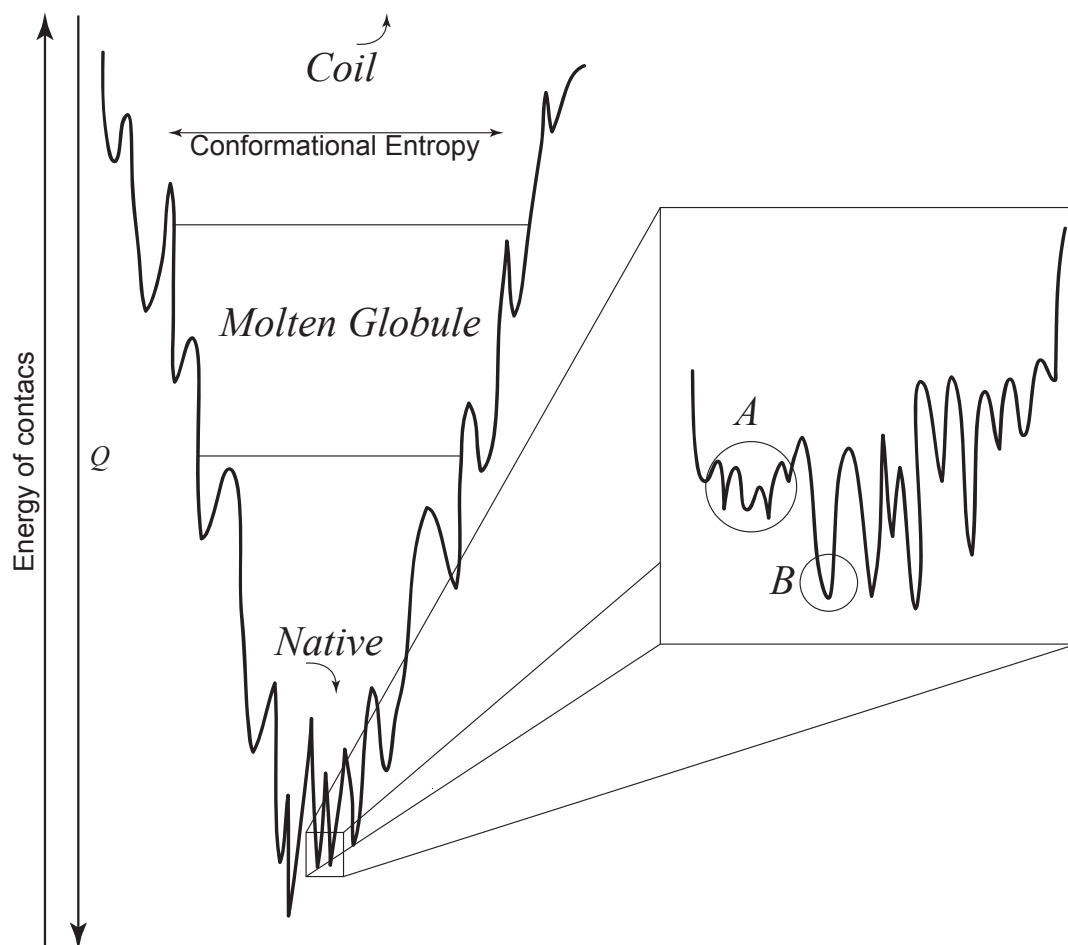
Figure 5.1: **The protein folding energy landscape is schematically shown in the shape of a funnel.**
The bottom of the funnel [204], containing the native, functional landscape, is zoomed in on the right. The new technique presented in this work allows to choose two conformations, *A* and *B*, from the native ensemble and calculate the free energy difference between them.
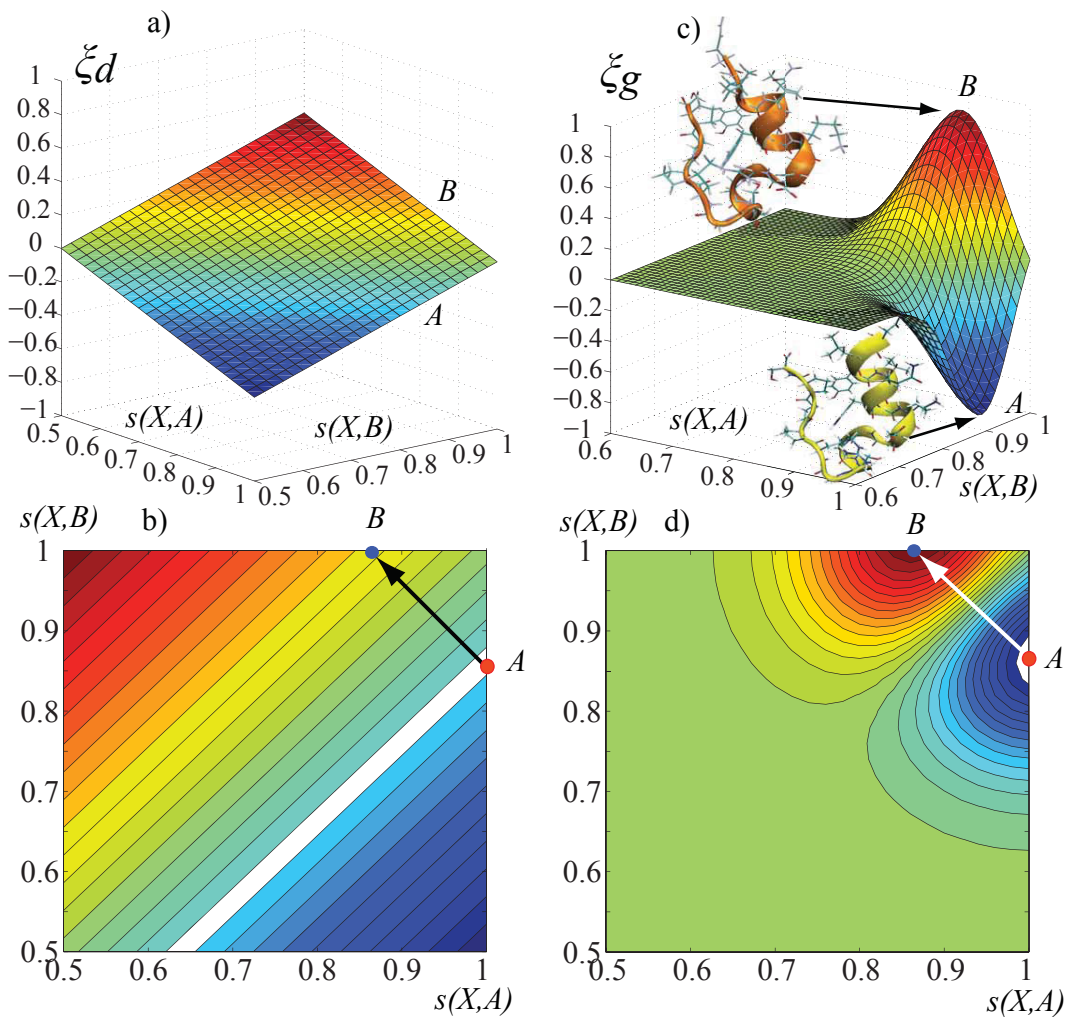
Figure 5.2: **Level lines for different path coordinates**

The path coordinate $\xi_d = s(X,A) - s(X,B)$ is shown in 3D (a) and as a contour plot (b). The path coordinate ξ_g (Eq.5.2) is shown in 3D (c) and as a contour plot (d). The labeled points correspond to conformations A and B – two of the states detected by an NMR study of Trp-cage native state [158]. A one-dimensional dynamical variable necessarily partitions the phase space into multidimensional iso-surfaces. In both cases on this figure iso-surfaces around $\xi_d \approx 0$ (a green stripe) and $\xi_g \approx 0$ (the large green area) contain all the unfolded states. This is the case with any one-dimensional path coordinate. Our method solves this problem with the confinement potential "tube".
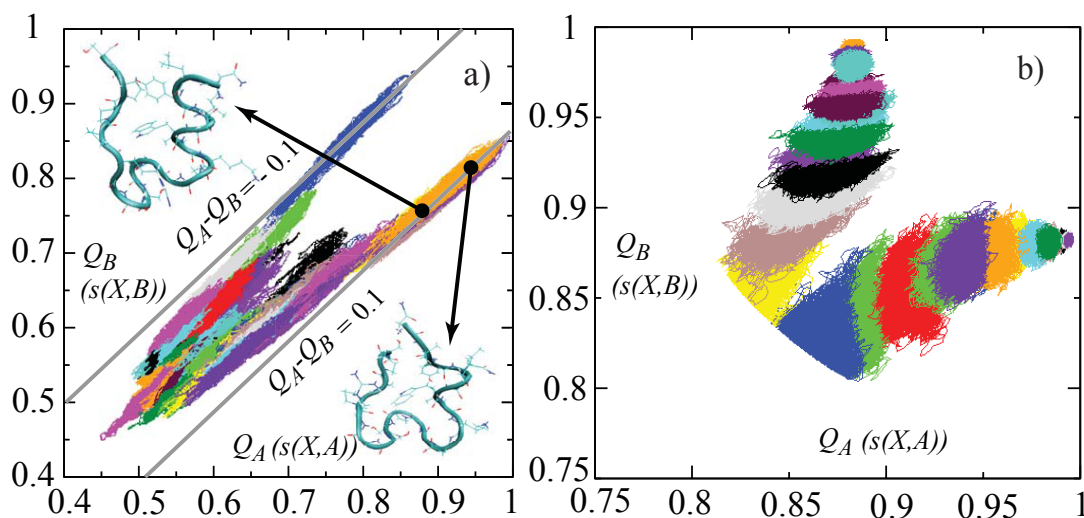
Figure 5.3: **Mapping of umbrella windows onto the phase space square.**
A patch of the same color corresponds to a trajectory in a single window. a) $\xi_d = s(X, A) - s(X, B) = Q_A - Q_B$ (Eq.5.1) performs poorly as a path coordinate. Two very dissimilar structures (shown in the corners) are in the same (orange) window, that also includes conformation A. b) ξ_g (Eq.5.2) performs much better with window patches covering the conformations A and B very compactly.
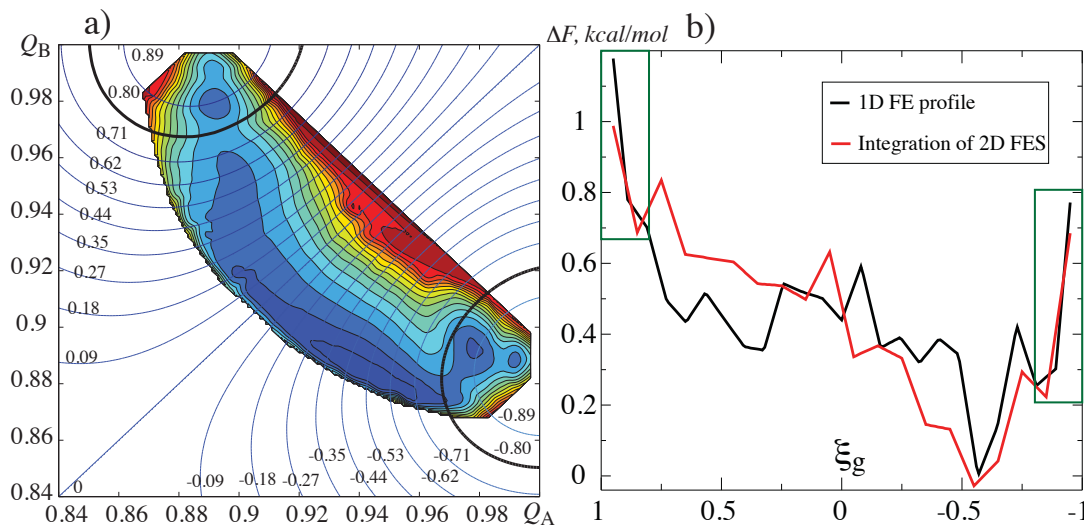
Figure 5.4: **Free energy maps.**
a) 2D free energy surface as a function of Q_A and Q_B; b) 1D free energy profiles obtained by integration of a FES in (a) and by umbrella sampling of ξ_g are very similar showing the consistency of the methods. In all simulations the confinement potential (5.5) was used. The basins of A and B are marked by thicker contour lines in (a) and by rectangles in (b).

Figure 5.5: **The confinement potential.**
The confinement phase space "tube" in our case is a wall of cylindrical shape, surrounding the native region of the phase space (corresponding to the upper right corner of the $((Q_A, Q_B)$ square)

Bibliography

[1] Mogilner, A & Oster, G. (1996) Cell motility driven by actin polymerization *Biophys. J.* **71**, 3030–45.
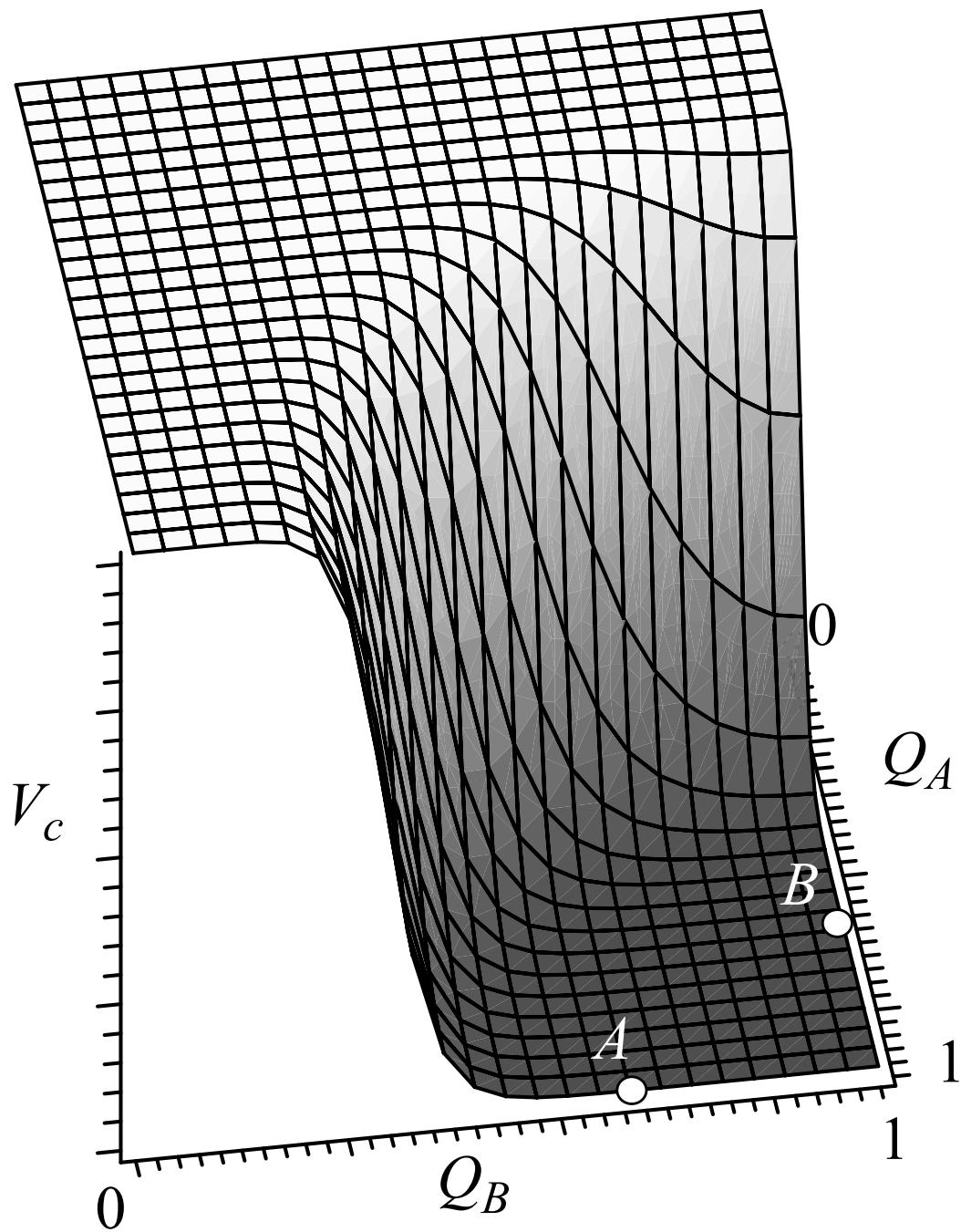
[2] Faix, J & Rottner, K. (2006) The making of filopodia. *Curr Opin Cell Biol* **18**, 18–25.

[3] Noselli, S. (2002) Drosophila, actin and videotape – new insights in wound healing. *Nat Cell Biol* **4**, E251–E253.

[4] Dent, E. W & Gertler, F. B. (2003) Cytoskeletal dynamics and transport in growth cone motility and axon guidance. *Neuron* **40**, 209–227.

[5] Jacinto, A, Wood, W, Balayo, T, Turmaine, M, Martinez-Arias, A, & Martin, P. (2000) Dynamic actin-based epithelial adhesion and cell matching during Drosophila dorsal closure *Curr. Biol.* **10**, 1420–1426.

[6] Lorenz, M, Yamaguchi, H, Wang, Y, Singer, R. H, & Condeelis, J. (2004) Imaging sites of N-wasp activity in lamellipodia and invadopodia of carcinoma cells. *Curr Biol* **14**, 697–703.

[7] Mejillano, M, Kojima, S.-I, Applewhite, D. A, Gertler, F. B, Svitkina, T. M, & Gary G Borisy, G. (2004) Lamellipodial versus filopodial mode of the actin nanomachinery: pivotal role of the filament barbed end *Cell* **118**, 363–373.

[8] Schaus, T. E, Taylor, E. W, & Gary G Borisy, G. (2007) Self-organization of actin filament orientation in the dendritic-nucleation/array-treadmilling model *Proc Natl Acad Sci USA* **104**, 7086–91.

[9] Cassimeris, L, Safer, D, Nachmias, V. T, & Zigmond, S. H. (1992) Thymosin beta 4 sequesters the majority of G-actin in resting human polymorphonuclear leukocytes *J Cell Biol* **119**, 1261–70.

[10] Vavylonis, D, Kovar, D. R, O'Shaughnessy, B, & Pollard, T. D. (2006) Model of formin-associated actin filament elongation. *Mol Cell* **21**, 455–466.

[11] Schirenbeck, A, Arasada, R, Bretschneider, T, Stradal, T. E. B, Schleicher, M, & Faix, J. (2006) The bundling activity of vasodilator-stimulated phosphoprotein is required for filopodium formation. *Proc Natl Acad Sci U S A* **103**, 7694–7699.

[12] Heid, P. J, Geiger, J, Wessels, D, Voss, E, & Soll, D. R. (2005) Computer-assisted analysis of filopod formation and the role of myosin II heavy chain phosphorylation in Dictyostelium. *J Cell Sci* **118**, 2225–2237.

[13] DiDonna, B. A & Levine, A. J. (2007) Unfolding cross-linkers as rheology regulators in F-actin networks *Phys Rev E Stat Nonlin Soft Matter Phys* **75**, 041909.

[14] Yamazaki, D, Kurisu, S, & Takenawa, T. (2005) Regulation of cancer cell motility through actin reorganization *Cancer Sci* **96**, 379–86.

[15] Lan, Y & Papoian, G. A. (2008) The stochastic dynamics of filopodial growth. *Biophys J* **94**, 3839–3852.

[16] Berg, J. S & Cheney, R. E. (2002) Myosin-X is an unconventional myosin that undergoes intrafilopodial motility. *Nat Cell Biol* **4**, 246–250.

[17] Tokuo, H & Ikebe, M. (2004) Myosin X transports Mena/VASP to the tip of filopodia. *Biochem Biophys Res Commun* **319**, 214–220.

[18] Bryngelson, J. D & Wolynes, P. G. (1987) Spin glasses and the statistical mechanics of protein folding. *Proc Natl Acad Sci U S A* **84**, 7524–7528.

[19] Onuchic, J. N, Wolynes, P. G, Luthey-Schulten, Z, & Socci, N. D. (1995) Toward an outline of the topography of a realistic protein-folding funnel. *Proc Natl Acad Sci U S A* **92**, 3626–3630.

[20] Frauenfelder, H, Sligar, S. G, & Wolynes, P. G. (1991) The energy landscapes and motions of proteins. *Science* **254**, 1598–1603.

[21] Frauenfelder, H, Alberding, N. A, Ansari, A, Braunstein, D, Cowen, B. R, Hong, M. K, Iben, I. E. T, Johnson, J. B, & Luck, S. (1990) Proteins and pressure *Journal of Physical Chemistry* **94**, 1024–1037.

[22] Shakhnovich, E. I & Gutin, A. M. (1993) Engineering of stable and fast-folding sequences of model proteins. *Proc Natl Acad Sci U S A* **90**, 7195–7199.

[23] Bryngelson, J. D, Onuchic, J. N, Socci, N. D, & Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195.

[24] Dobson, C. M, Sali, A, & Karplus, M. (1998) Protein Folding: A Perspective from Theory and Experiment *Angew. Chem. Int. Ed.* **37**, 868 – 893.

[25] Pande, V. S, Grosberg, A. Y, & Tanaka, T. (2000) Heteropolymer freezing and design: Towards physical models of protein folding *Rev Mod Phys* **72**, 259–314.

[26] Plotkin, S. S & Onuchic, J. N. (2002) Understanding protein folding with energy landscape theory. Part II: Quantitative aspects. *Q Rev Biophys* **35**, 205–286.

[27] Hyeon, C & Thirumalai, D. (2003) Can energy landscape roughness of proteins and RNA be measured by using mechanical unfolding experiments? *P Natl Acad Sci Usa* **100**, 10249–53.

[28] Scheraga, H. A, Khalili, M, & Liwo, A. (2007) Protein-folding dynamics: overview of molecular simulation techniques. *Annu Rev Phys Chem* **58**, 57–83.

[29] Dill, K. A, Ozkan, S. B, Shell, M. S, & Weikl, T. R. (2008) The protein folding problem. *Annu Rev Biophys* **37**, 289–316.

[30] Chen, Y, Ding, F, Nie, H, Serohijos, A. W, Sharma, S, Wilcox, K. C, Yin, S, & Dokholyan, N. V. (2008) Protein folding: then and now. *Arch Biochem Biophys* **469**, 4–19.

[31] McCammon, J. A, Gelin, B. R, & Karplus, M. (1977) Dynamics of folded proteins. *Nature* **267**, 585–590.

[32] Frauenfelder, H & McMahon, B. (1998) Dynamics and function of proteins: the search for general concepts. *Proc Natl Acad Sci U S A* **95**, 4795–4797.

[33] Henzler-Wildman, K & Kern, D. (2007) Dynamic personalities of proteins. *Nature* **450**, 964–972.

[34] Fenimore, P. W, Frauenfelder, H, McMahon, B. H, & Young, R. D. (2004) Bulk-solvent and hydration-shell fluctuations, similar to alpha- and beta-fluctuations in glasses, control protein motions and functions. *Proc Natl Acad Sci U S A* **101**, 14408–14413.

[35] Karplus, M & McCammon, J. A. (1981) THE INTERNAL DYNAMICS OF GLOBULAR-PROTEINS *Crc Cr Rev Bioch Mol* **9**, 293–349.

[36] Honeycutt, J. D & Thirumalai, D. (1992) The nature of folded states of globular proteins. *Biopolymers* **32**, 695–709.

[37] Bae, S.-H, Legname, G, Serban, A, Prusiner, S. B, Wright, P. E, & Dyson, H. J. (2009) Prion proteins with pathogenic and protective mutations show similar structure and dynamics *Biochemistry* **48**, 8120–8.

[38] Clore, G. M. (2008) Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement *Mol Biosyst* **4**, 1058–69.

[39] Tang, C, Louis, J. M, Aniana, A, Suh, J.-Y, & Clore, G. M. (2008) Visualizing transient events in amino-terminal autoprocessing of HIV-1 protease *Nature* **455**, 693–6.

[40] Thielges, M. C, Zimmermann, J, Yu, W, Oda, M, & Romesberg, F. E. (2008) Exploring the energy landscape of antibody-antigen complexes: protein dynamics, flexibility, and molecular recognition *Biochemistry* **47**, 7237–47.

[41] Fang, C, Frontiera, R. R, Tran, R, & Mathies, R. A. (2009) Mapping GFP structure evolution during proton transfer with femtosecond Raman spectroscopy *Nature* **462**, 200–4.

[42] Plotkin, S. S, Wang, J, & Wolynes, P. G. (1997) Statistical mechanics of a correlated energy landscape model for protein folding funnels *The Journal of Chemical Physics* **106**, 2932–2948.

[43] Derrida, B. (1985) A generalization of the random energy model which includes correlations between energies *J Phys Lett-Paris*.

[44] Guo, Z & Thirumalai, D. (1995) Kinetics of Protein Folding: Nucleation Mechanism, Time Scales, and Pathways *Biopolymers* **36**, 83 – 102.

[45] Gast, K, Damaschun, H, Misselwitz, R, Mller-Frohne, M, Zirwer, D, & Damaschun, G. (1994) Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. *Eur Biophys J* **23**, 297–305.

[46] Ptitsyn, O. B, Pain, R. H, Semisotnov, G. V, Zerovnik, E, & Razgulyaev, O. I. (1990) Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett* **262**, 20–24.

[47] Ptitsyn, O. B. (1996) How molten is the molten globule? *Nat Struct Biol* **3**, 488–490.

[48] Onuchic, J. N, Luthey-Schulten, Z, & Wolynes, P. G. (1997) Theory of protein folding: the energy landscape perspective *Annual Review of Physical Chemistry* **48**, 545–600.

[49] Chuang, J, Grosberg, A. Y, & Kardar, M. (2001) Free energy self-averaging in protein-sized random heteropolymers *Physical Review Letters* **87**, 078104.

[50] Geissler, P. L, Shakhnovich, E. I, & Grosberg, A. Y. (2004) Solvation versus freezing in a heteropolymer globule *Physical review E, Statistical, nonlinear, and soft matter physics* **70**, 021802.

[51] Hegler, J. A, Weinkam, P, & Wolynes, P. G. (2008) The spectrum of biomolecular states and motions *HFSP Journal* **2**, 307–13.

[52] Adcock, S. A & McCammon, J. A. (2006) Molecular dynamics: Survey of methods for simulating the activity of proteins *Chem Rev* **106**, 1589–1615.

[53] García, A. (1992) Large-amplitude nonlinear motions in proteins *Physical Review Letters* **68**, 2696–2699.

[54] Grant, B. J, Gorfe, A. A, & McCammon, J. A. (2009) Ras Conformational Switching: Simulating Nucleotide-Dependent Conformational Transitions with Accelerated Molecular Dynamics *PLoS Comput Biol* **5**, e1000325.

[55] Paschek, D, Hempel, S, & García, A. E. (2008) Computing the stability diagram Trp-cage miniprotein of the *P Natl Acad Sci Usa* **105**, 17754–17759.

[56] Cukier, R. I. (2009) Apo Adenylate Kinase Encodes Its Holo Form: A Principal Component and Varimax Analysis *J Phys Chem B* **113**, 1662–1672.

[57] Materese, C. K, Goldmon, C. C, & Papoian, G. A. (2008) Hierarchical organization of eglin c native state dynamics is shaped by competing direct and water-mediated interactions. *Proc Natl Acad Sci U S A* **105**, 10659–10664.

[58] Maisuradze, G. G, Liwo, A, & Scheraga, H. A. (2009) Principal Component Analysis for Protein Folding Dynamics *Journal of Molecular Biology* **385**, 312–329.

[59] Wood, W & Martin, P. (2002) Structures in focus–filopodia *Int J Biochem Cell Biol* **34**, 726–30.

[60] Miller, J, Fraser, S. E, & McClay, D. (1995) Dynamics of thin filopodia during sea urchin gastrulation *Development* **121**, 2501–11.

[61] Chang, S & Camilli, P. D. (2001) Glutamate regulates actin-based motility in axonal filopodia *Nat Neurosci* **4**, 787–93.

[62] Mogilner, A & Rubinstein, B. (2005) The physics of filopodial protrusion. *Biophys J* **89**, 782–795.

[63] Atilgan, E, Wirtz, D, & Sun, S. X. (2006) Membrane and dynamics of actin-driven thin membrane protrusions *Biophys. J.* **90**, 65.

[64] Schafer, D. A, Jennings, P. B, & Cooper, J. A. (1996) Dynamics of capping protein and actin assembly in vitro: uncapping barbed ends by polyphosphoinositides. *J Cell Biol* **135**, 169–179.

[65] Kepler, T. B & Elston, T. C. (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations *Biophys. J.* **81**, 3116–3136.

[66] Sasai, M & Wolynes, P. G. (2003) Stochastic gene expression as a many-body problem *Proc. Natl. Acad. Sci. USA* **100**, 2374–2379.

[67] Korobkova, E, Emonet, T, Vilar, J. M. G, Shimizu, T. S, & Cluzel, P. (2004) From molecular noise to behavioural variability in a single bacterium *Nature* **428**, 574.

[68] Walczak, A. M, Onuchic, J. N, & Wolynes, P. G. (2005) Absolute rate theories of epigenetic stability *Proc. Natl. Acad. Sci. USA* **102**, 18926.

[69] Weinberger, L. S, Burnett, J. C, Toettcher, J. E, Arkin, A. P, & Schaffer, D. V. (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuation drive phenotypic diversity *Cell* **122**, 169–182.

[70] Thattai, M & van Oudenaarden, A. (2004) Stochastic gene expressions in fluctuatiing environments *Genetics* **167**, 523–530.

[71] Lan, Y, Wolynes, P. G, & Papoian, G. A. (2006) A variational approach to the stochastic aspects of cellular signal transduction *J. Chem. Phys* **125**, 124106.

[72] Lan, Y & Papoian, G. A. (2006) The interplay between discrete noise and nonlinear chemical kinetics in a signal amplification cascade *J. Chem. Phys* **125**, 154901.

[73] Lan, Y & Papoian, G. A. (2007) Stochastic resonant signaling in enzyme cascades *Phys Rev Lett* **98**, 228301.

[74] Gomez, T. M, Robles, E, Poo, M, & Spitzer, N. C. (2001) Filopodial calcium transients promote substrate-dependent growth cone turning. *Science* **291**, 1983–1987. Several minutes.

[75] Jontes, J. D, Buchanan, J, & Smith, S. J. (2000) Growth cone and dendrite dynamics in zebrafish embryos: early events in synaptogenesis imaged in vivo. *Nat Neurosci* **3**, 231–237. exponential distribution with 3 mins constant.

[76] McCroskery, S, Chaudhry, A, Lin, L, & Daniels, M. P. (2006) Transmembrane agrin regulates filopodia in rat hippocampal neurons in culture. *Mol Cell Neurosci* **33**, 15–28.

[77] Portera-Cailliau, C, Pan, D. T, & Yuste, R. (2003) Activity-regulated dynamic behavior of early dendritic protrusions: evidence for different types of dendritic filopodia. *J Neurosci* **23**, 7129–7142. 3-7 mins.

[78] Varnum-Finney, B & Reichardt, L. F. (1994) Vinculin-deficient PC12 cell lines extend unstable lamellipodia and filopodia and have a reduced rate of neurite outgrowth. *J Cell Biol* **127**, 1071–1084. Several minutes.

[79] Pronk, S, Geissler, P. L, & Fletcher, D. A. (2008) Limits of filopodium stability. *Phys Rev Lett* **100**, 258102.

[80] Artyomov, M. N, Das, J, Kardar, M, & Chakraborty, A. K. (2007) Purely stochastic binary decisions in cell signaling models without underlying deterministic bistabilities *P Natl Acad Sci Usa* **104**, 18958–63.

[81] Zong, C, Lu, T, Shen, T, & Wolynes, P. G. (2006) Nonequilibrium self-assembly of linear fibers: microscopic treatment of growth, decay, catastrophe and rescue *Phys Biol* **3**, 83–92.

[82] Kuramoto, Y. (1984) *Chemical Oscillations, Waves and Turbulence*.

[83] Gillespie, D. T. (1977) Exact Stochastic Simulation of coupled chemical reactions *J. Phys. Chem.* **81**, 2340–2361.

[84] Svitkina, T. M, Bulanova, E. A, Chaga, O. Y, Vignjevic, D. M, Kojima, S, Vasiliev, J. M, & Borisy, G. G. (2003) Mechanism of filopodia initiation by reorganization of a dendritic network *J. Cell Biol.* **160**, 409–421.

[85] Argiro, V, Bunge, M. B, & Johnson, M. I. (1985) A quantitative study of growth cone filopodial extension. *J Neurosci Res* **13**, 149–162.

[86] Gustafson, T & Wolpert, L. (1961) Studies on the cellular basis of morphogenesis in the sea urchin embryo: directed movements of primary mesenchyme cells in normal and vegetalized larvae *Exp. Cell Res.* **24**, 64–79.

[87] Pantaloni, D, Clainche, C. L, & Carlier, M. F. (2001) Mechanism of actin-based motility. *Science* **292**, 1502–1506.

[88] Pollard, T. D & Borisy, G. G. (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**, 453–465.

[89] Carlsson, A. E. (2001) Growth of branched actin networks against obstacles *Biophysical Journal* **81**, 1907–23.

[90] Mogilner, A. (2006) On the edge: modeling protrusion *Curr Opin Cell Biol* **18**, 32–9. References: [25] [27].

[91] Mogilner, A & Oster, G. (2003) Force generation by actin polymerization II: the elastic ratchet and tethered filaments *Biophys. J.* **84**, 1591–605.

[92] Peskin, C. S, Odell, G. M, & Oster, G. F. (1993) Cellular motions and thermal fluctuations: the Brownian ratchet *Biophys. J.* **65**, 316–324.

[93] Lin, L. C.-L & Brown, F. L. H. (2004) Brownian dynamics in Fourier space: membrane simulations over long length and time scales. *Phys Rev Lett* **93**, 256001.

[94] Pécréaux, J, Döbereiner, H.-G, Prost, J, Joanny, J.-F, & Bassereau, P. (2004) Refined contour analysis of giant unilamellar vesicles *The European physical journal E, Soft matter* **13**, 277–90.

[95] Gov, N. S & Safran, S. A. (2005) Red blood cell membrane fluctuations and shape controlled by ATP-induced cytoskeletal defects *Biophys. J.* **88**, 1859–74.

[96] Safran, S. A, Gov, N. S, Nicolas, A, Schwarz, U, & Tlusty, T. (2005) Physics of cell elasticity, shape and adhesion *Physica a-Statistical Mechanics and Its Applications* **352**, 171–201. 927KR Times Cited:2 Cited References Count:88.

[97] Mallavarapu, A & Mitchison, T. J. (1999) Regulated actin cytoskeleton assembly at filopodium tips controls their extension and retraction *J. Cell Biol.* **146**, 1097.

[98] Lidke, D. S, Lidke, K. A, Rieger, B, Jovin, T. M, & Arndt-Jovin, D. J. (2005) Reaching out for signals: filopodia sense EGF and respond by directed retrograde transport of activated receptors. *J Cell Biol* **170**, 619–626.

[99] Forscher, P & Smith, S. J. (1988) Actions of cytochalasins on the organization of actin filaments and microtubules in a neuronal growth cone *J Cell Biol* **107**, 1505–16.

[100] Brown, M & Bridgman, P. (2003) Retrograde flow rate is increased in growth cones from myosin IIB knockout mice *Journal of cell science* **116**, 1087–1094.

[101] Huckaba, T. M, Lipkin, T, & Pon, L. A. (2006) Roles of type II myosin and a tropomyosin isoform in retrograde actin flow in budding yeast *J Cell Biol* **175**, 957–69.

111

[102] Lawson, N. D & Weinstein, B. M. (2002) In vivo imaging of embryonic vascular development using transgenic zebrafish. *Dev Biol* **248**, 307–318.

[103] Kovar, D. R. (2007) Intracellular motility: myosin and tropomyosin in actin cable flow. *Curr Biol* **17**, R244–R247.

[104] Lin, C. H, Espreafico, E. M, Mooseker, M. S, & Forscher, P. (1996) Myosin drives retrograde F-actin flow in neuronal growth cones. *Neuron* **16**, 769–782.

[105] Zhuravlev, P. I & Papoian, G. A. (2009) Molecular noise of capping protein binding induces macroscopic instability in filopodial dynamics. *Proc Natl Acad Sci U S A* **106**, 11570–5.

[106] Mellor, H. (2009) The role of formins in filopodia formation *Biochim. Biophys. Acta*.

[107] Naoz, M, Manor, U, Sakaguchi, H, Kachar, B, & Gov, N. S. (2008) Protein localization by actin tread-milling and molecular motors regulates stereocilia shape and treadmilling rate. *Biophys J* **95**, 5706–5718.

[108] Prost, J, Barbetta, C, & Joanny, J.-F. (2007) Dynamical control of the shape and size of stereocilia and microvilli. *Biophys J* **93**, 1124–1133.

[109] Gov, N. S. (2006) Dynamics and morphology of microvilli driven by actin polymerization. *Phys Rev Lett* **97**, 018101.

[110] Rifkin, J. L & Speisman, R. A. (1976) Filamentous extensions of vegetative amoebae of the cellular slime mold Dictyostelium. *Trans Am Microsc Soc* **95**, 165–173.

[111] Berg, J. S, Powell, B. C, & Cheney, R. E. (2001) A millennial myosin census *Mol Biol Cell* **12**, 780–94.

[112] Pollard, T. D, Blanchoin, L, & Mullins, R. D. (2000) Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu Rev Biophys Biomol Struct* **29**, 545–576.

[113] Drenckhahn, D & Pollard, T. D. (1986) Elongation of actin filaments is a diffusion-limited reaction at the barbed end and is accelerated by inert macromolecules. *J Biol Chem* **261**, 12754–12758.

[114] Zicha, D, Dobbie, I. M, Holt, M. R, Monypenny, J, Soong, D. Y. H, Gray, C, & Dunn, G. A. (2003) Rapid actin transport during cell protrusion. *Science* **300**, 142–145.

[115] Berg, J. S, Derfler, B. H, Pennisi, C. M, Corey, D. P, & Cheney, R. E. (2000) Myosin-X, a novel myosin with pleckstrin homology domains, associates with regions of dynamic actin. *J Cell Sci* **113 (19)**, 3439–3451.

[116] Bohil, A. B, Robertson, B. W, & Cheney, R. E. (2006) Myosin-X is a molecular motor that functions in filopodia formation. *Proc Natl Acad Sci U S A* **103**, 12411–12416.

[117] Sheetz, M. P, Wayne, D. B, & Pearlman, A. L. (1992) Extension of filopodia by motor-dependent actin assembly. *Cell Motil Cytoskeleton* **22**, 160–169.

[118] Kerber, M. L, Jacobs, D. T, Campagnola, L, Dunn, B. D, Yin, T, Sousa, A. D, Quintero, O. A, & Cheney, R. E. (2009) A Novel Form of Motility in Filopodia Revealed by Imaging Myosin-X at the Single-Molecule Level *Current Biology* **19**, 967 – 973.

[119] Zhang, H, Berg, J. S, Li, Z, Wang, Y, Lng, P, Sousa, A. D, Bhaskar, A, Cheney, R. E, & Strłmblad, S. (2004) Myosin-X provides a motor-based link between integrins and the cytoskeleton. *Nat Cell Biol* **6**, 523–531.

[120] Applewhite, D. A, Barzik, M, Kojima, S.-I, Svitkina, T. M, Gertler, F. B, & Borisy, G. G. (2007) Ena/VASP proteins have an anti-capping independent function in filopodia formation. *Mol Biol Cell* **18**, 2579–2591.

[121] Barzik, M, Kotova, T. I, Higgs, H. N, Hazelwood, L, Hanein, D, Gertler, F. B, & Schafer, D. A. (2005) Ena/VASP proteins enhance actin polymerization in the presence of barbed end capping proteins. *J Biol Chem* **280**, 28653–28662.

[122] Krause, M, Dent, E. W, Bear, J. E, Loureiro, J. J, & Gertler, F. B. (2003) Ena/VASP proteins: regulators of the actin cytoskeleton and cell migration. *Annu Rev Cell Dev Biol* **19**, 541–564.

[123] Hüttelmaier, S, Harbeck, B, Steffens, O, Messerschmidt, T, Illenberger, S, & Jockusch, B. M. (1999) Characterization of the actin binding properties of the vasodilator-stimulated phosphoprotein VASP. *FEBS Lett* **451**, 68–74.

[124] Chereau, D & Dominguez, R. (2006) Understanding the role of the G-actin-binding domain of Ena/VASP in actin assembly. *J Struct Biol* **155**, 195–201.

[125] Knight, P. J, Thirumurugan, K, Xu, Y, Wang, F, Kalverda, A. P, Stafford, W. F, Sellers, J. R, & Peckham, M. (2005) The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head. *J Biol Chem* **280**, 34702–34708.

[126] Nagy, S, Ricca, B. L, Norstrom, M. F, Courson, D. S, Brawley, C. M, Smithback, P. A, & Rock, R. S. (2008) A myosin motor that selects bundled actin for motility. *Proc Natl Acad Sci U S A* **105**, 9616–9620.

[127] Koestler, S. A, Rottner, K, Lai, F, Block, J, Vinzenz, M, & Small, J. V. (2009) F- and G-actin concentrations in lamellipodia of moving cells *PLoS ONE* **4**, e4810.

[128] Pollard, T. D. (1986) Rate constants for the reactions of ATP- and ADP-actin with the ends of actin filaments *J Cell Biol* **103**, 2747–54.

[129] Weber, A, Nachmias, V. T, Pennise, C. R, Pring, M, & Safer, D. (1992) Interaction of thymosin beta 4 with muscle and platelet actin: implications for actin sequestration in resting platelets *Biochemistry* **31**, 6179–85.

[130] Novak, I. L, Slepchenko, B. M, & Mogilner, A. (2008) Quantitative analysis of G-actin transport in motile cells *Biophysical Journal* **95**, 1627–38.

[131] Cramer, L. P, Briggs, L. J, & Dawe, H. R. (2002) Use of fluorescently labelled deoxyribonuclease I to spatially measure G-actin levels in migrating and non-migrating cells *Cell Motil Cytoskeleton* **51**, 27–38.

[132] Salles, F. T, Merritt, R. C, Manor, U, Dougherty, G. W, Sousa, A. D, Moore, J. E, Yengo, C. M, Dosé, A. C, & Kachar, B. (2009) Myosin IIIa boosts elongation of stereocilia by transporting espin 1 to the plus ends of actin filaments *Nat Cell Biol* **11**, 443–50.

[133] Belyantseva, I. A, Boger, E. T, & Friedman, T. B. (2003) Myosin XVa localizes to the tips of inner ear sensory cell stereocilia and is essential for staircase formation of the hair bundle *P Natl Acad Sci Usa* **100**, 13958–63.

[134] Delprat, B, Michel, V, Goodyear, R, Yamasaki, Y, Michalski, N, El-Amraoui, A, Perfettini, I, Legrain, P, Richardson, G, Hardelin, J.-P, & Petit, C. (2005) Myosin XVa and whirlin, two deafness gene products required for hair bundle growth, are located at the stereocilia tips and interact directly *Hum Mol Genet* **14**, 401–10.

[135] Sekerková, G, Zheng, L, Loomis, P. A, Changyaleket, B, Whitlon, D. S, Mugnaini, E, & Bartles, J. R. (2004) Espins are multifunctional actin cytoskeletal regulatory proteins in the microvilli of chemosensory and mechanosensory cells *J Neurosci* **24**, 5445–56.

[136] McConnell, R. E & Tyska, M. J. (2007) Myosin-1a powers the sliding of apical membrane along microvillar actin bundles *J Cell Biol* **177**, 671–81.

[137] El-Amraoui, A, BAHLOUL, A, & Petit, C. (2007) Myosin VII *Proteins And Cell Regulation* **Volume 7**, 353–373.

[138] Wicki, A, Lehembre, F, Wick, N, Hantusch, B, Kerjaschki, D, & Christofori, G. (2006) Tumor invasion in the absence of epithelial-mesenchymal transition: podoplanin-mediated remodeling of the actin cytoskeleton. *Cancer Cell* **9**, 261–272.

[139] Stukalin, E. B & Kolomeisky, A. B. (2006) ATP hydrolysis stimulates large length fluctuations in single actin filaments. *Biophys J* **90**, 2673–2685.

[140] Raucher, D & Sheetz, M. P. (2000) Cell spreading and lamellipodial extension rate is regulated by membrane tension. *J Cell Biol* **148**, 127–136.

[141] Kern, D & Zuiderweg, E. R. P. (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* **13**, 748–757.

[142] Clarkson, M. W, Gilmore, S. A, Edgell, M. H, & Lee, A. L. (2006) Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry* **45**, 7693–7699.

[143] Helmstaedt, K, Krappmann, S, & Braus, G. H. (2001) Allosteric regulation of catalytic activity: Escherichia coli aspartate transcarbamoylase versus yeast chorismate mutase. *Microbiol Mol Biol Rev* **65**, 404–21, table of contents.

[144] Hong, M. K, Braunstein, D, Cowen, B. R, Frauenfelder, H, Iben, I. E, Mourant, J. R, Ormos, P, Scholl, R, Schulte, A, & Steinbach, P. J. (1990) Conformational substates and motions in myoglobin. External influences on structure and dynamics. *Biophys J* **58**, 429–436.

[145] Thorpe, I. F & Brooks, C. L. (2007) Molecular evolution of affinity and flexibility in the immune system *P Natl Acad Sci Usa* **104**, 8821–6.

[146] Dioumaev, A. K & Lanyi, J. K. (2007) Bacteriorhodopsin photocycle at cryogenic temperatures reveals distributed barriers of conformational substates *Proceedings of the National Academy of Sciences* **104**, 9621–9626.

[147] Brooks, C. L, Onuchic, J. N, & Wales, D. J. (2001) Statistical Thermodynamics: Taking a Walk on a Landscape *Science* **293**, 612–613.

[148] Frauenfelder, H & Wolynes, P. G. (1994) Biomolecules: Where the Physics of Complexity and Simplicity Meet *Physics Today* **47**, 58–64.

[149] Gruebele, M. (2002) Protein folding: the free energy surface *Current Opinion in Structural Biology* **12**, 161–168.

[150] Wales, D. J, Miller, M. A, & Walsh, T. R. (1998) Archetypal energy landscapes *Nature* **394**, 758–760.

[151] Becker, O. M & Karplus, M. (1997) The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics *The Journal of Chemical Physics* **106**, 1495–1517.

[152] Levy, Y & Becker, O. M. (2001) Energy landscapes of conformationally constrained peptides *Journal of Chemical Physics* **114**, 993–1009.

[153] Krivov, S. V & Karplus, M. (2004) Hidden complexity of free energy surfaces for peptide (protein) folding. *Proc Natl Acad Sci U S A* **101**, 14766–14770.

[154] Metzler, R, Klafter, J, & Jortner, J. (1999) Hierarchies and logarithmic oscillations in the temporal relaxation patterns of proteins and other complex systems. *Proc Natl Acad Sci U S A* **96**, 11085–11089.

[155] Miyashita, O, Onuchic, J. N, & Wolynes, P. G. (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins *Proc Natl Acad Sci USA* **100**, 12570–5.

[156] Simmerling, C, Strockbine, B, & Roitberg, A. E. (2002) All-atom structure prediction and folding simulations of a stable protein. *J Am Chem Soc* **124**, 11258–11259.

[157] Snow, C, Zagrovic, B, & Pande, V. (2002) The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations *Journal of the American Chemical Society* **124**, 14548–14549.

[158] Neidigh, J. W, Fesinmeyer, R. M, & Andersen, N. H. (2002) Designing a 20-residue protein. *Nat Struct Biol* **9**, 425–430.

[159] Plotkin, Wang, & Wolynes. (1996) Correlated energy landscape model for finite, random heteropolymers. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **53**, 6271–6296.

[160] Hardin, C, Eastwood, M. P, Luthey-Schulten, Z, & Wolynes, P. G. (2000) Associative memory hamiltonians for structure prediction without homology: alpha-helical proteins. *Proc Natl Acad Sci U S A* **97**, 14235–14240.

[161] Socci, N. D, Onuchic, J. N, & Wolynes, P. G. (1996) Diffusive dynamics of the reaction coordinate for protein folding funnels *The Journal of Chemical Physics* **104**, 5860–5868.

[162] Bursulaya, B & Brooks, C. (1999) Folding Free Energy Surface of a Three-Stranded Beta-Sheet Protein *Journal of the American Chemical Society* **121**, 9947–9951.

[163] Zhou, R. (2003) Free energy landscape of protein folding in water: explicit vs. implicit solvent. *Proteins* **53**, 148–161.

[164] MacKerell, A. D, Banavali, N, & Foloppe, N. (2000) Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257–265.

[165] Kumar, S, Rosenberg, J. M, Bouzida, D, Swendsen, R. H, & Kollman, P. A. (1992) The weighted histogram analysis method for free-energy calculations on biomolecules *Journal of Computational Chemistry* **13(8)**, 1011 – 1021.

[166] Kohlbacher, O & Lenhof, H. P. (2000) BALL–rapid software prototyping in computational molecular biology. Biochemicals Algorithms Library. *Bioinformatics* **16**, 815–824.

[167] Plimpton, S. (1995) Fast Parallel Algorithms for Short-Range Molecular Dynamics *Journal of Computational Physics* **117**, 1–19.

[168] Phillips, J. C, Braun, R, Wang, W, Gumbart, J, Tajkhorshid, E, Villa, E, Chipot, C, Skeel, R. D, Kalé, L, & Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781–1802.

[169] Garcia-Ojalvo, J & M.Sancho, J. (1999) *Noise in Spatially Extended Systems*. (Springer-Verlag New York, Inc).

[170] Ward, J. H. (1963) Hierarchical Grouping to Optimize an Objective Function *Journal of the American Statistical Association,* **58**, 236–244.

[171] Zhou, R. (2003) Trp-cage: folding free energy landscape in explicit water. *Proc Natl Acad Sci U S A* **100**, 13280–13285.

[172] Gallivan, J. P & Dougherty, D. A. (1999) Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A* **96**, 9459–9464.

[173] Shehu, A, Kavraki, L. E, & Clementi, C. (2007) On the Characterization of Protein Native State Ensembles *Biophys. J.* **92**, 1503–1511.

[174] Shenu, A, Clementi, C, & Kavraki, L. E. (2007) Sampling Conformation Space to Model Equilibrium Fluctuations in Proteins *Algorithmica* **48**, 303–327.

[175] Chen, J, Dima, R. I, & Thirumalai, D. (2007) Allosteric communication in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J Mol Biol* **374**, 250–266.

[176] Bahar, I & Rader, A. J. (2005) Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol* **15**, 586–592.

[177] Hayward, S & Go, N. (1995) Collective Variable Description of Native Protein Dynamics *Annual Review of Physical Chemistry* **46**, 223–250.

[178] Ma, J. (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. *Structure* **13**, 373–380.

[179] Caflisch, A. (2006) Network and graph analyses of folding free energy surfaces. *Curr Opin Struct Biol* **16**, 71–78.

[180] Schwille, P, Kummer, S, Heikal, A. A, Moerner, W. E, & Webb, W. W. (2000) Fluorescence correlation spectroscopy reveals fast optical excitation-driven intramolecular dynamics of yellow fluorescent proteins. *Proc Natl Acad Sci U S A* **97**, 151–156.

[181] Yang, W. Y & Gruebele, M. (2003) Folding at the speed limit. *Nature* **423**, 193–197.

[182] Chekmarev, S. F, Krivov, S. V, & Karplus, M. (2006) Folding of ubiquitin: a simple model describes the strange kinetics. *J Phys Chem B* **110**, 8865–8869.

[183] Liu, F, Du, D, Fuller, A. A, Davoren, J. E, Wipf, P, Kelly, J. W, & Gruebele, M. (2008) An experimental survey of the transition between two-state and downhill protein folding scenarios. *Proc Natl Acad Sci U S A* **105**, 2369–2374.

[184] Juraszek, J & Bolhuis, P. G. (2006) Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc Natl Acad Sci U S A* **103**, 15859–15864.

[185] Zhuravlev, P. I & Papoian, G. A. (2010) Functional versus folding landscapes: the same yet different *Current Opinion in Structural Biology* **20**, 16–20.

[186] Henzler-Wildman, K. A, Lei, M, Thai, V, Kerns, S. J, Karplus, M, & Kern, D. (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916.

[187] Frauenfelder, H, Parak, F. G, & Young, R. D. (1988) Conformational substates in proteins *Annual review of biophysics and biophysical chemistry* **17**, 451–79.

[188] Levy, Y, Cho, S. S, Onuchic, J. N, & Wolynes, P. G. (2005) A survey of flexible protein binding mechanisms and their transition states using native topology based energy landscapes. *J Mol Biol* **346**, 1121–1145.

[189] Zhuravlev, P. I, Materese, C. K, & Papoian, G. A. (2009) Deconstructing the native state: energy landscapes, function, and dynamics of globular proteins *The Journal of Physical Chemistry B* **113**, 8800–12.

[190] Hyeon, C & Onuchic, J. N. (2007) Mechanical control of the directional stepping dynamics of the kinesin motor *P Natl Acad Sci Usa* **104**, 17382–17387.

[191] Kollman, P. A, Massova, I, Reyes, C, Kuhn, B, Huo, S, Chong, L, Lee, M, Lee, T, Duan, Y, Wang, W, Donini, O, Cieplak, P, Srinivasan, J, Case, D. A, & Cheatham, T. E. (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models *Acc Chem Res* **33**, 889–97.

[192] Weis, A, Katebzadeh, K, Söderhjelm, P, Nilsson, I, & Ryde, U. (2006) Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field *J Med Chem* **49**, 6596–606.

[193] Savelyev, A & Papoian, G. A. (2007) Inter-DNA electrostatics from explicit solvent molecular dynamics simulations *J Am Chem Soc* **129**, 6060–1.

[194] Vorobjev, Y. N & Hermans, J. (1999) ES/IS: estimation of conformational free energy by combining dynamics simulations with explicit solvent with an implicit solvent continuum model *Biophys Chem* **78**, 195–205.

[195] Park, S, Lau, A. Y, & Roux, B. (2008) Computing conformational free energy by deactivated morphing *The Journal of Chemical Physics* **129**, 134102.

[196] Freddolino, P. L, Park, S, Roux, B, & Schulten, K. (2009) Force field bias in protein folding simulations *Biophysical Journal* **96**, 3772–80.

[197] Banavali, N. K & Roux, B. (2005) Free energy landscape of A-DNA to B-DNA conversion in aqueous solution. *J Am Chem Soc* **127**, 6866–6876.

[198] Torrie, G & Valleau, J. (1977) Non-physical Sampling Distributions In Monte-Carlo Free-Energy Estimation- Umbrella Sampling *Journal of Computational Physics* **23**, 187–199.

[199] Whitford, P. C, Gosavi, S, & Onuchic, J. N. (2008) Conformational transitions in adenylate kinase - Allosteric communication reduces misligation *J Biol Chem* **283**, 2042–2048.

[200] Pisliakov, A, Cao, J, Kamerlin, S, & Warshel, A. (2009) Enzyme millisecond conformational dynamics do not catalyze the chemical step *P Natl Acad Sci Usa*.

[201] Wu, S, Zhuravlev, P. I, & Papoian, G. A. (2008) High resolution approach to the native state ensemble kinetics and thermodynamics *Biophysical Journal* **95**, 5524–32.

[202] Sugita, Y & Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding *Chemical Physics Letters* **314**, 141–151.

[203] Ferrenberg, A. M & Swendsen, R. H. (1989) Optimized Monte Carlo data analysis *Phys. Rev. Lett.* **63**, 1195–1198.

[204] Onuchic, J. N & Wolynes, P. G. (2004) Theory of protein folding *Current Opinion in Structural Biology* **14**, 70–5.