2018

CONFORMATIONS, ENERGETICS, AND KINETICS OF PEPTIDES IN MEMBRANE ENVIRONMENT

Ganga P. Sharma
University of Rhode Island, gp_sharma@uri.edu

CONFORMATIONS, ENERGETICS, AND KINETICS OF PEPTIDES IN MEMBRANE ENVIRONMENT

BY

GANGA P. SHARMA

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHYSICS

UNIVERSITY OF RHODE ISLAND

2018

DOCTOR OF PHILOSOPHY DISSERTATION

OF

GANGA P. SHARMA

APPROVED:

Dissertation Committee:

Major Professor Gerhard Müller

Yana Reshetnyak

David Freeman

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2018

ABSTRACT

Chapter1 : We present the exact solution of a microscopic statistical mechanical model for the transformation of a long polypeptide between an unstructured coil conformation and an α-helix conformation. The polypeptide is assumed to be adsorbed to the interface between a polar and a non-polar environment such as realized by water and the lipid bilayer of a membrane. The interfacial coil-helix transformation is the first stage in the folding process of helical membrane proteins. Depending on the values of model parameters, the conformation changes as a crossover, a discontinuous transition, or a continuous transition with helicity in the role of order parameter. Our model is constructed as a system of statistically interacting quasiparticles that are activated from the helix pseudo-vacuum. The particles represent links between adjacent residues in coil conformation that form a self-avoiding random walk in two dimensions. Explicit results are presented for helicity, entropy, heat capacity, and the average numbers and sizes of both coil and helix segments.

Chapter 2: We investigate profiles of local attributes (densities of entropy, enthalpy, free energy, and helicity) for the backbone of long polypeptides in the heterogeneous environment of a lipid bilayer or cell membrane. From these profiles we infer landscapes of global attributes for the backbone of short peptides with given position and orientation in that environment. Our methodology interprets the broken internal H-bonds along the backbone of the polypeptide as statistically interacting quasiparticles activated from the helix reference state. The interaction depends on the local environment (ranging from polar to non-polar), in particular on the availability of external H-bonds (with H_2O molecules or lipid headgroups) to replace internal H-bonds. The helicity landscape in particular is an essential prerequisite for the continuation of this part of the project with focus on the side-

chain contributions to the free-energy landscapes. The full free-energy landscapes are expected to yield information on insertion conditions and likely insertion pathways.

Chapter 3: We present the first part in the design of a kinetic model for the insertion of short peptides, including variants of pHLIP, into a lipid bilayer. The process under scrutiny combines a transport phenomenon and a change in protonation status of negatively charged sites near the C terminus. The two kinetic phenomena influence each other and set different time scales. Processes with a significant range of time scales, known to be a challenge for molecular dynamics simulations, are shown to be within the scope of the kinetic modeling presented here, which is based on interlocking Markov chain processes. The two processes governing protonation status and transport are run individually and then in combination. This makes it possible to investigate feedback mechanisms between the two component processes.

ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor Professor Gerhard Müller. It has been an honor to be his Ph.D. student. Prof. Müller provided me with every bit of guidance, assistance, and expertise throughout my studies.

I would like to thank our experimental collaborators Professors Yana Reshetnyak and Oleg Andreev of URI for their valuable suggestions in all the projects that we have collaborated. Specifically, I would like to thank Prof. Reshetnyak for her continued support and for providing me the access of her Laboratory. I would like to thank my friend, Slaybaugh Gregory, for his great support and help during my quest of learning biophysical experiments. I would also like to thank my colleague, Aaron Mayer, for important discussions and suggestions.

I also extend my gratitude to Professor David Freeman for his great support and invaluable advice. I am thankful to Prof. Leonard M. Kahn, Graduate Program Director, for his help, kindness, and guidance during my Ph.D. years. I am deeply indebted to him. I am also grateful to Prof. David Chelidze for his support, insightful comments and time.

I would especially like to express my deepest gratitude to all my family members and friends. This dissertation would not have been possible without their warm love, continued patience, and endless support. More specifically, I would like to thank and dedicate this dissertation to my my late Mom, Nirjala Sharma, and Dad, Janak P. Sharma, for their love and inspiration throughout my life. I must express my gratitude to my wife, Gita, for her continued support throughout this experience. To my beloved daughter Garima and son Gaurav, I would like to express my thanks for being good kids always cheering me up. Finally, I would like to thank to my brother, Parashuram Upadhyaya, who originally ignited my passion for science and thirst for knowledge.

DEDICATION

In loving memory of my mother, Nirjala Sharma(1949-2010)

PREFACE

This dissertation develops a theoretical model for the membrane-associated insertion/folding of short peptides, employing advanced concepts of statistical physics and kinetic theory. My work has been motivated by experiments carried out in the Biological Physics Laboratory of Professors Yana Reshetnyak and Oleg Andreev at URI. The three chapters represent integral parts of the modeling framework with focus on conformational change, energetics, and kinetics, respectively. Each of the three chapters has the format of a research paper and therefore manuscript format is in use. The first paper has already been published. The second and third papers are in preparation for publication.

1. G. P. Sharma, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, *Coil-helix transition of polypeptide at water-lipid interface*, Journal of Statistical Mechanics: Theory and Experiment, **P01034**, pp.1-23 (2015).

2. G. P. Sharma, A. C. Meyer, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, *Free-energy landscapes for peptides in a membrane environment I: Contributions of the backbone* (manuscript in preparation of publication).

3. G. P. Sharma, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, *Kinetic model for peptide insertion into a membrane I: kinetics of a protonable residue at varying pH* (manuscript in preparation of publication).

TABLE OF CONTENTS

Page

Page

CHAPTER 1

Coil-Helix Transition of Peptides at Water-Lipid Interface

1.1 Introduction

The folding mechanisms of water-soluble proteins from primary to secondary and higher-order structures has been thoroughly investigated over many years. In the study of protein translocation pathways into and across cell membranes, which is a very important active area of current research, one important problem requiring further elucidation is the coil-helix transition that accompanies the insertion of a polypeptide into a lipid bilayer. The theoretical modeling of this ubiquitous process in biological matter is fairly complex due to the heterogeneous environment in which conformational changes occur and the simultaneity or rapid succession of conformational change and translocation. Experimental studies are limited to the small selection of polypeptides that are water soluble and undergo controllable insertion/folding and exit/unfolding processes.

The folding of all helical membrane proteins/peptides, independent of the insertion mechanism, is governed by the formation of helical segments in the lipid bilayer environment. This process is driven by hydrophobic interactions and hydrogen bonding [1, 2, 3, 4]. Its two main steps are the transformation from coil to interfacial helix and the insertion of the helix into the membrane with transmembrane orientation. Variants of the *pH Low Insertion Peptide* (pHLIP) family are water soluble and prove to be well suited for the investigation of membrane-associated folding and unfolding [5, 6], reversibly driven by changes in pH. A drop in pH leads to the protonation of negatively charged side chains, which enhances the hydrophobicity of the peptide and initiates the aforementioned two-step process of folding and insertion. A subsequent rise in pH reverses the process: the peptide unfolds and exits. Recent experimental studies have already established

important thermodynamic and kinetic parameters of the peptide-membrane interaction [7, 8, 9].

What has been lacking for these and related experiments is a microscopic statistical mechanical model with experimentally testable attributes that is amenable to an exact analysis. Our goal is to construct, solve, develop, and test such a model in three successive stages. The first stage, which is the theme of this paper, involves the design and solution of a microscopic model that describes the coil-helix transformation of a long polypeptide adsorbed to the lipid bilayer of a membrane (see Fig. 1.1). Such a model is an indispensable part of a theory of membrane-

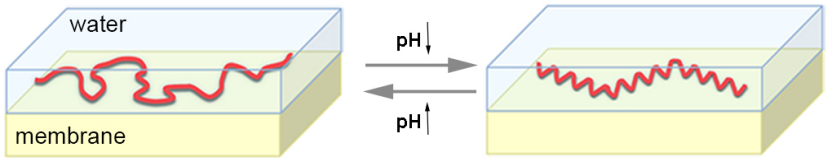

Figure 1.1. (Color online) Long polypeptide at the interface between water and a flat lipid bilayer undergoing a reversible and pH-driven coil -helix transformation.

associated folding and will be used as the foundation for the next two stages. They include (a) the investigation of profiles of local attributes for generic polypeptides and landscapes of global attributes for short peptides such as pHLIP in the heterogeneous water/lipid environment and (b) the kinetics of insertion and exit as can be inferred from the landscapes of free energy and conformational attributes.

The pioneering theoretical studies of coil-helix transformations and related phenomena that appeared throughout 1960s were admirably compiled and reviewed in a monograph by Poland and Scheraga [10]. A series of model systems were introduced at that time. Many of them are still being used today in textbooks and research papers. This includes the familiar Zimm-Bragg model [11] and generalizations thereof, all amenable to the transfer matrix method of analysis. Also emerging at that time was the highly original Lifson method of statistical me-

chanical analysis [12], the scope of which includes (smooth) crossovers and (sharp) transitions [13, 14]. The need at this time for yet another model solved by yet a different method is dictated by the three stages of our project as will become apparent in what follows.

In Sec. 1.2 we present the microscopic model of our design as a system of statistically interacting links and describe the method of its exact analysis. Depending on the parameter settings the exact solution produces a conformational change in the form of a crossover or a transition (Sec. 1.3). The transition may be of first or second order as we discuss in Sec. 1.4 with focus on the helicity (order parameter) and entropy (measure of disorder) and other quantities. In Sec. 1.5 we summarize the main advances of this work and point out their role as the foundation for the continuation of this project in two different directions in the arenas of biological physics and statistical mechanics. We also discuss how this work connects to recent studies by other researchers and how the continuation of this project can benefit from those studies.

1.2 Model system

The microscopic model that we present here is a system of statistically interacting quasiparticles with shapes. The methodology employed for its exact statistical mechanical analysis is built on the concept of fractional statistics, invented by Haldane [15], and developed by Wu [16], Isakov [17], Anghel [18], and others [19] in the context of quantum many-body systems. The adaptation of this approach to classical statistical mechanical systems of particles with shapes was developed in a recent series of studies with applications to Ising spins [20, 21, 22, 23], jammed granular matter [24, 25], lattice gases with long-range interactions [26], and DNA under tension [27]. The application to the coil-helix transition of a long polypeptide adsorbed to a water-lipid interface worked out in the following is conceptually

simple but surprisingly rich in scope.

1.2.1 Coil segments from helix vacuum

The reference state (pseudo-vacuum) of our model system is the ordered helix conformation of N residues bound by peptide bonds into $N-1$ links and stabilized by internal hydrogen bonds along the backbone. Thermal fluctuations or environmental change cause the nucleation of disordered coil segments, which then grow by unravelling adjacent helical order.

In our model the coil segments are represented by thermally activated links that combine to form a self-avoiding random walk between the ends of successive helical segments. Both coil and helix segments are confined to the water-lipid interface (Fig. 1.1). Coil segments carry configurational entropy that grows with their size and range in the interfacial plane. That range is controllable at a microscopic level by the integer-valued model parameter μ, henceforth called range parameter.

Each residue can be in $\mu + 1$ states, of which one (denoted h) represents the helix conformation and μ (numbered $\mathsf{1}$, $\ldots$, μ) represent the coil conformation. Access by a residue to these states is constrained by the states of its neighboring residues as illustrated in Fig. 1.2 for the case $\mu = 3$.

Helical links (hh) are short and form straight segments (horizontal, in Fig. 1.2) in the plane of the water-lipid interface. Coil links are more extended and directed either horizontally ($\mathsf{11}$, $\mathsf{22}$, $\mathsf{33}$) or vertically ($\mathsf{h1}$, $\mathsf{12}$, $\mathsf{23}$, $\mathsf{32}$, $\mathsf{21}$, $\mathsf{1h}$). Our model allows each coil segment to randomly explore the interface on one side of the helical direction without intersecting itself.[1] That space is discretized and constrained by the length of the segment and by the number of distinct coil states.

We assign different activation energies to coil links relative to hh links depend-

[1] The one-side restriction has no significant impact on the quantities we are calculating in this work. Generalizations to coil segments that explore the plane of the interface more freely are planned.

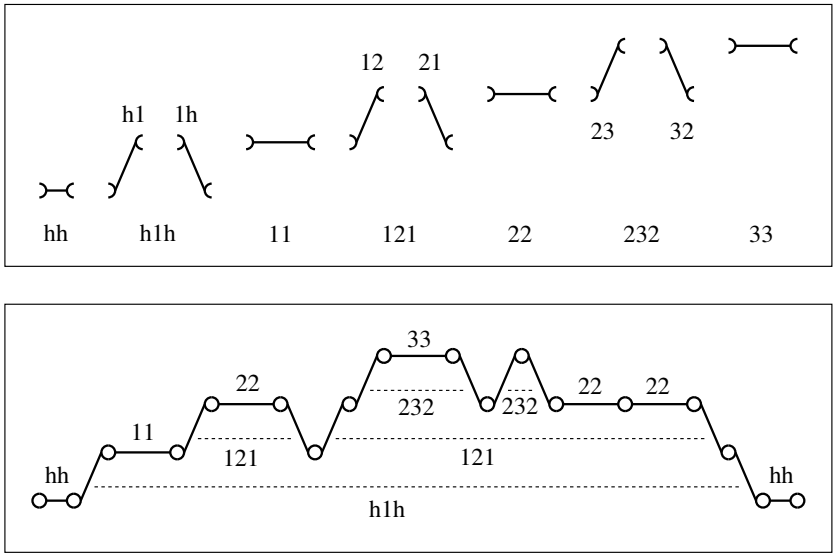

Figure 1.2. Segment of coil conformation between two segments of helix conformation (bottom), generated by the activation of 2μ species of statistically interacting particles in the form of single links or composed of pairs of links that are not necessarily adjacent (top). Residues in helix conformation are in a unique state (h). Residues in coil conformation are in one of μ states $(1, 2, ..., \mu)$, constrained to form a self-avoiding walk in the interfacial plane and here illustrated for $\mu = 3$.

ing on whether they contribute to nucleation or to growth. To the former we assign the activation energy $\epsilon_{\rm n}$ and to the latter $\epsilon_{\rm g}$ (two model parameters). Nucleation of one coil link requires the simultaneous break-up of several internal hydrogen bonds whereas growth proceeds by the break-up of one bond per link (two bonds shared by different pairs of residues). On our way to calculating a partition function we now face the task of counting microstates of given link content.

1.2.2 Combinatorics of links

For the combinatorial analysis we introduce a set of statistically interacting quasiparticles that contain individual links or pairs of links. It turns out that we need 2μ species of particles. In the case of $\mu = 3$ they are the six species (along with the element of pseudo-vacuum) shown in the top panel of Fig. 1.2. The combinatorics of statistically interacting particles is captured by the energy

expression

$$E(\{N_m\}) = E_{\rm pv} + \sum_{m=1}^{2\mu} N_m \epsilon_m, \tag{1.1}$$

and the multiplicity expression [15, 16, 17, 18, 19, 20, 21],

$$W(\{N_m\}) = \prod_{m=1}^{2\mu} \left(\begin{array}{c} d_m + N_m - 1 \\ N_m \end{array} \right), \tag{1.2}$$

$$d_m = A_m - \sum_{m'=1}^{2\mu} g_{mm'}(N_{m'} - \delta_{mm'}), \tag{1.3}$$

inferred from a generalized Pauli principle as will be illustrated below for the application under consideration. This means that there exist $W(\{N_m\})$ microstates with energy $E(\{N_m\})$ that all have the same particle content: N_m particles of species m for $m = 1, \ldots, 2\mu$. The ϵ_m are particle activation energies relative to the energy $E_{\rm pv}$ of the pseudo-vacuum, the A_m are capacity constants, and the $g_{mm'}$ are statistical interaction coefficients. The specifications for the model with $\mu = 3$ distinct coil states are compiled in Table 1.1.

Table 1.1. Specifications of the six species of particles that describe the case $\mu = 3$.

motif	cat.	m	ϵ_m	A_m
h1h	host	1	$\epsilon_{\rm n}$	$N-2$
121	hybrid	2	$2\epsilon_{\rm g}$	0
232	hybrid	3	$2\epsilon_{\rm g}$	0
11	tag	4	$\epsilon_{\rm g}$	0
22	tag	5	$\epsilon_{\rm g}$	0
33	tag	6	$\epsilon_{\rm g}$	0

$g_{mm'}$	1	2	3	4	5	6
1	2	2	2	1	1	1
2	−1	0	0	0	0	0
3	0	−1	0	0	0	0
4	−1	−1	0	0	0	0
5	0	−1	−1	0	0	0
6	0	0	−1	0	0	0

The particles form nested structures as indicated in Fig. 1.2. We have species from three categories in the taxonomy of Ref. [21]: one species of *hosts*, two species of *hybrids*, and three species of *tags*. Hosts cannot be hosted, tags cannot host any particles from a different species, hybrids can do both.

A system of N residues in the helix pseudo-vacuum has the capacity of nucleating a coil segment at $A_1 = N - 2$ different locations by activating a host particle

with activation energy $\epsilon_1 = \epsilon_{\rm n}$. The activation of particles from any species reduces the capacity d_1 of the system for placing further hosts on account of (1.3) and $g_{1m} > 0$. Hosts and hybrids have twice the size of tags. The former thus reduce the capacity at double the rate of the latter.

For any mix of particles the system has a finite capacity. When that capacity has been reached, we have $d_1 = 1$, which makes the associated binomial factor in (1.2) equal to one. If we attempt to add one more particle from any species, d_1 becomes zero or a negative integer and, in consequence, the associated binomial factor vanishes. The helix pseudo-vacuum has zero capacity for the placement of hybrids and tags, as implied by $A_2 = \cdots = A_6 = 0$. Such capacity of the system is generated dynamically by the placement of particles with hosting capacity. This generation of capacity is encoded in the negative interaction coefficients. Hosts 1 generate capacity for placing hybrids 2 and tags 4. Hybrids 2, in turn, generate capacity for placing hybrids 3 and tags, 4, 5 etc. Tags do not generate capacity for placing any particles.

The particle content of the coil segment of $N = 16$ residues shown in Fig. 1.2 is $N_1 = 1$, $N_2 = 2$, $N_3 = 2$, $N_4 = 1$, $N_5 = 3$, $N_6 = 1$. Its activation energy (1.1) thus becomes $E(1,2,2,1,3,1) - E_{\rm pv} = \epsilon_{\rm n} + 13\epsilon_{\rm g}$ and the number of coil segments of equal contour length and with the same particle content is, according to (1.2), (1.3), $W(1,2,2,1,3,1) = 360$. Further microstates with equal activation energy are generated if we exchange hybrids or tags from one species by hybrids or tags from a different species or if we replace hybrids by pairs of tags (or vice versa), all within the constraints imposed by the nesting. The constraints are encoded in the multiplicity expression. It does not allow spurious particle combinations.

The generalization to any μ is straightforward: the zoo of 2μ particle species now comprises one host, $\mu - 1$ hybrids, and μ tags, labeled consecutively in this

order. The host has activation energy $\epsilon_{\rm n}$, reflecting the nucleation of coil segments, whereas hybrids and tags have activation energies $2\epsilon_{\rm g}$ and $\epsilon_{\rm g}$, respectively, reflecting the growth of coil segments. The capacity constants remain the same for each category: $A_m = (N-2)\delta_{m,1}$.

The nonzero interaction coefficients generalize naturally in accordance with the sizes and nested structure of the particles: $g_{1m} = 2$ (1) for $m = 1, \ldots, \mu$ ($\mu + 1, \ldots, 2\mu$); $g_{m'm} = -1$ for three sets of index pairs: (i) $m = m' - 1$, $m' = 2, \ldots, \mu$; (ii) $m = m' - \mu$, $m' = \mu+1, \ldots, 2\mu$; (iii) $m = m' - \mu + 1$, $m' = \mu+1, \ldots, 2\mu-1$. The case $\mu = 1$ has no hybrids: $g_{11} = 2$, $g_{12} = 1$, $g_{21} = -1$, $g_{22} = 0$. It is equivalent to the Zimm-Bragg model [11]. With the combinatorial analysis completed we turn to the statistical mechanical analysis.

1.2.3 Free energy of polypeptide

The partition function for the adsorbed polypeptide, modeled as a system of statistically interacting and thermally activated particles [15, 16, 17, 18, 19, 20, 21],

$$Z = \sum_{\{N_m\}} W(\{N_m\}) e^{-\beta E(\{N_m\})}, \tag{1.4}$$

depends on energy (1.1) and multiplicity (1.2) with ingredients $\epsilon_m, A_m, g_{mm'}$ from Sec. 1.2.2. The thermal equilibrium macrostate in the thermodynamic limit follows from an extremum principle. Its implementation yields the partition function for a macroscopic system in the (generic) form [16, 17, 18, 19, 20, 26],

$$Z = \prod_{m=1}^{M} \left(1 + w_m^{-1}\right)^{A_m}, \tag{1.5}$$

where $M = 2\mu$ in our case and the (real, positive) w_m are solutions of the coupled nonlinear equations,

$$e^{\beta\epsilon_m} = (1 + w_m) \prod_{m'=1}^{M} \left(1 + w_{m'}^{-1}\right)^{-g_{m'm}}. \tag{1.6}$$

The average number of particles from species m are derived from the coupled linear equations,

$$w_m N_m + \sum_{m'=1}^{M} g_{mm'} N_{m'} = A_m. \tag{1.7}$$

It is useful and economical to express all results as functions of the two control parameters[2]

$$\tau \doteq e^{\beta(\epsilon_{\mathrm{g}} - \epsilon_{\mathrm{n}})} \qquad : 0 \leq \tau \leq 1, \tag{1.8}$$

$$t \doteq e^{\beta \epsilon_{\mathrm{g}}} \qquad : 0 \leq t < \infty, \tag{1.9}$$

with an additional dependence on the discrete range parameter μ implied. The nucleation parameter τ is a measure of cooperativity and controls the average length of coil and helix segments. High cooperativity ($\tau \ll 1$) means a high nucleation threshold. Low cooperativity ($\tau \lesssim 1$) means little difference in enthalpic cost of nucleation and growth. The growth parameter t controls the preference of one or the other conformation. Coil is preferred at small t and helix at large t.

Equations (1.6) for $w_1, \ldots, w_{2\mu}$ with parameters t, τ used on the left and the $g_{mm'}$ from Sec. 1.2.2 used on the right can be reduced to a single polynomial equation of order $\mu + 1$ for $w_{\mu+1}$:

$$(1 + w_{\mu+1} - t) S_\mu(w_{\mu+1}) = t\tau S_{\mu-1}(w_{\mu+1}), \tag{1.10}$$

where the $S_m(w)$ are Chebyshev polynomials of the second kind. Among all the solutions of Eq. (1.10) there exists exactly one,

$$w \doteq w_{\mu+1}(t, \tau), \tag{1.11}$$

that satisfies the criterion of physical relevance, requiring that (1.11) and all the

[2]The relevant energy scales are the strength of hydrogen bonds (~ 5 kcal/mol) and $k_{\mathrm{B}}T$ at room temperature (~ 0.6 kcal/mol).

remaining w_m inferred from it via

$$w_1 = \frac{S_\mu(w)}{\tau S_{\mu-1}(w)} = \frac{t}{1+w-t},$$
$$w_m = \begin{cases} \dfrac{S_{\mu-m+2}(w)}{S_{\mu-m}(w)} & : m = 2, \ldots, \mu, \\ w & : m = \mu+1, \ldots, 2\mu, \end{cases} \tag{1.12}$$

are non-negative. The derivation of this reduction is outlined in Appendix A.

The Gibbs free energy per residue inferred from (1.5) then depends on that physical solution as follows:

$$\bar{G}(t,\tau) = -k_{\mathrm{B}}T \ln\left(1 + w_1^{-1}\right), \tag{1.13}$$

from which any thermodynamic quantity of interest can be derived, including the entropy,

$$\bar{S} \doteq -\left(\frac{\partial \bar{G}}{\partial T}\right)_{\epsilon_{\mathrm{n}},\epsilon_{\mathrm{g}}}, \tag{1.14}$$

the enthalpy,

$$\bar{H} \doteq \bar{G} + T\bar{S}, \tag{1.15}$$

the helicity (order parameter),

$$\bar{N}_{\mathrm{hl}} \doteq 1 - \left(\frac{\partial \bar{G}}{\partial \epsilon_{\mathrm{n}}}\right)_{T,\epsilon_{\mathrm{g}}} - \left(\frac{\partial \bar{G}}{\partial \epsilon_{\mathrm{g}}}\right)_{T,\epsilon_{\mathrm{n}}}, \tag{1.16}$$

the density of (helix or coil) segments,

$$\bar{N}_{\mathrm{seg}} \doteq \left(\frac{\partial \bar{G}}{\partial \epsilon_{\mathrm{n}}}\right)_{T,\epsilon_{\mathrm{g}}}, \tag{1.17}$$

and the average sizes of helix segments and coil segments,

$$L_{\mathrm{hs}} \doteq \frac{\bar{N}_{\mathrm{hl}}}{\bar{N}_{\mathrm{seg}}}, \quad L_{\mathrm{cs}} \doteq \frac{1-\bar{N}_{\mathrm{hl}}}{\bar{N}_{\mathrm{seg}}}. \tag{1.18}$$

The population densities $\bar{N}_m \doteq N_m/N$, $m = 1, \ldots, 2\mu$, of particles can be extracted from the solution of the linear Eqs. (1.7) as shown in Appendix A.

1.3 Structure of solution

Changing the level of pH primarily affects the growth parameter t. At normal pH we have $t \lesssim 1$, which favors the random coil conformation. A drop in pH pushes the growth parameter to higher values, $t > 1$, which increasingly favors a conformation with helical ordering.[3] Depending on the value of the nucleation parameter τ and the discrete parameter μ, which controls the amount of entropy that coil segments can generate, the growth of helicity takes place in a crossover or in a transition of first or second order. To illuminate the criteria for these alternatives we investigate the nature of the physically relevant solution (1.11) of Eq. (1.10), in particular the singularities it acquires in the limits $\tau \to 0$ at $\mu < \infty$ and $\mu \to \infty$ at $\tau > 0$.

1.3.1 Crossover

For $\tau > 0$ and $\mu < \infty$ the solution $w(t, \tau)$ is bounded from below by

$$w_0 \doteq 2 \cos\left(\frac{\pi}{\mu + 1}\right), \tag{1.19}$$

which is the location of the last zero of $S_\mu(w)$. That value is only realized at $t = 0$ as illustrated in Fig. 1.3(a). For $t \gg 1$ the solution converges toward the linear asymptote,

$$w_{\text{as}} \doteq t + \tau - 1. \tag{1.20}$$

Note that w_0 depends on μ but not on τ whereas w_{as} depends on τ but not on μ. The smooth dependence on t of $w(t, \tau)$ for $\tau > 0$ and $\mu < \infty$ describes a crossover from low helicity at small t to high helicity at large t.

The Zimm-Bragg model [11] is represented by the case $\mu = 1$, for which (1.10)

[3] The level of pH effectively controls how easy or hard it is to replace broken internal hydrogen bonds along the backbone of the polypeptide with external hydrogen bonds involving H_2O molecules. Hence the shift in t.

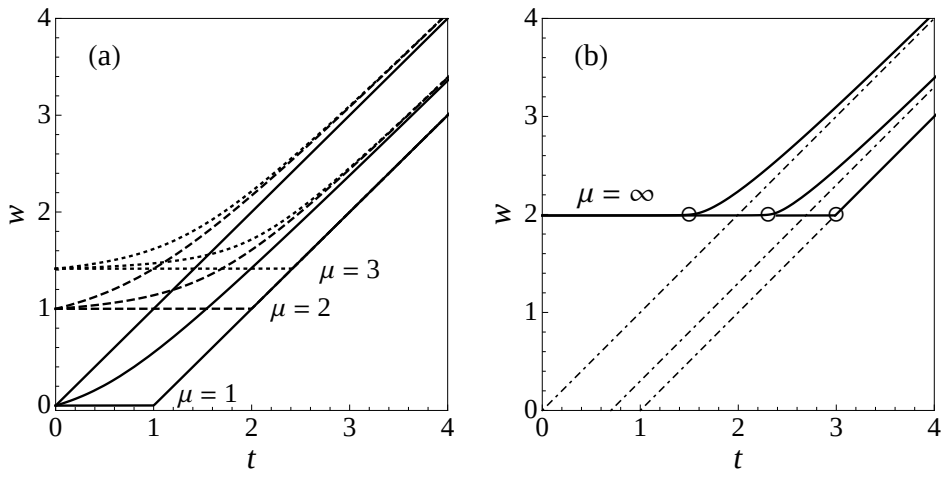

Figure 1.3. Physical solution (1.11) of the polynomial equation (1.10) for $\tau = 1, 0.3, 0$ (left to right in both panels). Panel (a) shows the emergence of a singularity in the limit $\tau \to 0$ for $\mu = 1, 2, 3$ and panel (b) the emergence of a singularity in the limit $\mu \to \infty$ at $\tau \geq 0$.

is a quadratic equation with physical solution

$$w = \frac{1}{2}\left[t - 1 + \sqrt{(t-1)^2 + 4t\tau}\,\right]. \tag{1.21}$$

It is mathematically equivalent to an Ising chain.[4]

The case $\mu = 2$ is qualitatively different in that each coil segment now carries entropy. The physical solution of the associated (cubic) Eq. (1.10) reads

$$w = \frac{1}{3}\left[x + 2\sqrt{x^2 + 3y}\cos\frac{\varphi}{3}\right], \tag{1.22}$$
$$\tan\varphi = \frac{\sqrt{27(4y^3 + y^2x^2 + 18yx^2 + 4x^4 - 27x^2)}}{x(2x^2 + 9y - 27)},$$

where $x \doteq t - 1$, $y \doteq 1 + t\tau$, and $0 \leq \varphi < \pi$.

For $3 \leq \mu < \infty$ and $\tau > 0$ the solution (1.11) must be determined numerically. In that context it is advisable to rewrite (1.10) as

$$(w + 1 - t)r_\mu(w) - t\tau = 0, \quad r_\mu(w) \doteq \frac{S_\mu(w)}{S_{\mu-1}(w)}, \tag{1.23}$$

[4]The Zimm-Bragg parameters commonly used are $\sigma = \tau$ and $s = t$. The (physically relevant) larger eigenvalue of the transfer matrix is $\lambda_0 = w + 1$, taking into account a shift in energy scale by ϵ_g per residue.

using the function

$$r_\mu(w) = \begin{cases} \dfrac{1}{2}\left[w + \sqrt{4 - w^2}\right] \cot\left(\mu \arccos\dfrac{w}{2}\right) & : w < 2, \\ \dfrac{\mu + 1}{\mu} & : w = 2, \\ \dfrac{1}{2}\left[w + \sqrt{w^2 - 4}\right] \coth\left(\mu \operatorname{arcosh}\dfrac{w}{2}\right) & : w > 2 \end{cases} \tag{1.24}$$

inferred from trigonometric/hyperbolic representations of Chebyshev polynomials. For large μ the functions $r_\mu(w)$ are much smoother than the polynomials $S_\mu(w)$. Standard methods with initial values from the analytic solution for $\mu \to \infty$ derived in Sec. 1.3.3 below work quite well.

1.3.2 First-order transition

In the limit $\tau \to 0$ at $\mu < \infty$, the solution (1.11) acquires a linear cusp as shown in Fig. 1.3(a):

$$w = \begin{cases} t_0 - 1 & : t \le t_0 \\ t - 1 & : t \ge t_0 \end{cases} \qquad (\tau = 0), \tag{1.25}$$

as the growth parameter t increases across the transition value,

$$t_0 \doteq 1 + 2\cos\left(\frac{\pi}{\mu + 1}\right) \qquad (\tau = 0). \tag{1.26}$$

It describes a discontinuous phase transition between a pure coil at $t < t_0$ and a pure helix at $t > t_0$ in a sense that requires some explanations (Sec. 1.4). Discontinuities are manifest in the order parameter and the entropy. The latter is associated with a latent heat.

1.3.3 Second-order transition

In the limit $\mu \to \infty$ at $\tau > 0$ the solution (1.11) acquires a quadratic cusp at

$$t_\mathrm{c} \doteq \frac{3}{1 + \tau} \qquad (\mu = \infty). \tag{1.27}$$

Performing that limit in (1.24) yields

$$r_\infty(w) = \frac{1}{2}\left[w + \sqrt{w^2 - 4}\right], \quad w \ge 2. \tag{1.28}$$

The resulting analytic solution then reads

$$w = \begin{cases} 2 & : 0 \le t \le t_{\rm c} \\ t - 1 + \dfrac{t\tau}{\lambda} & : t > t_{\rm c} \end{cases} \qquad (\mu = \infty), \tag{1.29}$$

$$\lambda \doteq \frac{1}{2}\left[t - 1 + \sqrt{(t+1)(t-3) + 4t\tau}\,\right], \tag{1.30}$$

and is graphically represented in Fig. 1.3(b). The singularity at $t_{\rm c}^{+}$,

$$w = 2 + \left[\frac{t_0(t - t_{\rm c})}{t_{\rm c}(t_0 - t_{\rm c})}\right]^2 + \mathcal{O}\big((t - t_{\rm c})^3\big) \tag{1.31}$$

with $t_0 = 3$ for $\mu = \infty$ represents a continuous transition between a coil phase ($t < t_{\rm c}$) and a helix phase ($t > t_{\rm c}$) with helical ordering subject to thermal fluctuations. Expression (1.31) is to be interpreted as an asymptotic expansion with coefficients that diverge as $\tau \to 0$. In that limit the cusp turns linear as in (1.25).

1.4 Order and disorder

Helix means order and coil means disorder, clearly. However, both attributes can be looked at from different angles and a more comprehensive picture emerges. In the following we investigate several thermodynamic quantities, derived from the free energy (1.13) as functions of the experimentally controllable growth parameter t at fixed values of the other two parameters τ and μ.

Each quantity will illuminate the competition between order and disorder from a somewhat different vantage point. All are functions of $w_1(t, \tau)$, which depends on the solution (1.11) of (1.10) via (1.12). The analytic expression in the limit $\tau \to 0$ for $\mu < \infty$ as inferred from (1.25) reads

$$w_1 = \begin{cases} \dfrac{t}{t_0 - t} & : t < t_0 \\ \infty & : t \ge t_0 \end{cases} \qquad (\tau = 0), \tag{1.32}$$

and the analytic result in the limit $\mu \to \infty$ at $\tau > 0$ as inferred from (1.29) becomes

$$w_1 = \begin{cases} \dfrac{t}{t_0 - t} & : t < t_{\rm c} \\ \dfrac{\lambda}{\tau} & : t \ge t_{\rm c} \end{cases} \qquad (\mu = \infty). \tag{1.33}$$

1.4.1 Helicity and entropy

We begin by considering the two thermodynamic functions that represent order and disorder most directly: helicity (1.16),

$$\bar{N}_{\rm hl} = 1 - \frac{t}{w_1(1+w_1)}\frac{\partial w_1}{\partial t}, \tag{1.34}$$

and entropy (1.14),

$$\frac{\bar{S}}{k_{\rm B}} = \ln\left(1+w_1^{-1}\right) + \frac{1}{w_1(1+w_1)}\left[t\ln t\frac{\partial w_1}{\partial t} + \tau\ln\tau\frac{\partial w_1}{\partial \tau}\right]. \tag{1.35}$$

In Figs. 1.4 and 1.5 we show the dependence of helicity and entropy on the growth parameter t at fixed nucleation parameter τ (five curves) and range parameter μ (two panels).

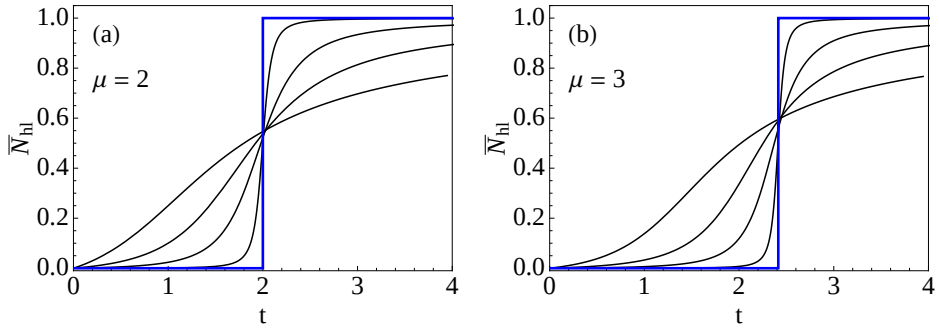

Figure 1.4. (Color online) Helicity $\bar{N}_{\rm hl}$ versus growth parameter t at values $\tau = 1, 0.25, 0.05, 0.0025$ (thin lines) and $\tau = 0$ (thick line) of the nucleation parameter for (a) $\mu = 2$ and (b) $\mu = 3$.

At finite cooperativity ($\tau > 0$) the helicity crosses over from a low to a high value near t_0. The rise in helicity becomes sharper with increasing cooperativity and turns into a step discontinuity in the limit $\tau \to 0$. Analytically, expression (1.34) with (1.32) substituted yields

$$\bar{N}_{\rm hl} = \theta(t - t_0) \qquad (\tau = 0). \tag{1.36}$$

While order as reflected in the helicity increases monotonically with t, the disorder as reflected in the entropy is not monotonically decreasing. It shows a

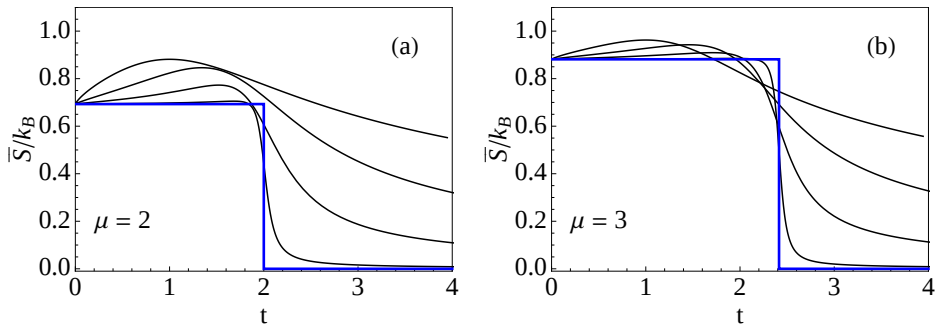

Figure 1.5. (Color online) Entropy $\bar{S}/k_{\rm B}$ versus t at $\tau = 1, 0.25, 0.05, 0.0025$ (thin lines) and $\tau = 0$ (thick line) for (a) $\mu = 2$ and (b) $\mu = 3$.

shallow maximum at $t \simeq 1$ separate from the shoulder at $t \simeq t_0$. The reason for this difference is that there is only one source of order – helical links – but two sources of disorder: disorder inside coil segments and disorder in the sequence of helical/coil segments of diverse lengths. It is the first source of disorder that produces the shoulder and the second source that produces the shallow maximum. The Zimm-Bragg case $\mu = 1$ is pathological in this respect. It produces coil segments without internal entropy as noted and commented on before [28].

As $\tau \to 0$, the segments grow larger in size and become fewer in numbers (see Sec. 1.4.2 below). This reduces disorder of the second kind. At infinite cooperativity the entropy turns into a step discontinuity of μ-dependent height and location. Expression (1.35) with (1.32) substituted becomes

$$\frac{\bar{S}}{k_B} = \theta(t_0 - t) \ln t_0 \qquad (\tau = 0). \tag{1.37}$$

This discontinuity signals the presence of a latent heat (see Sec. 1.4.3 below).

Next we investigate the same measures of order and disorder as functions of t at fixed $\tau = 1.0$ (low cooperativity) or $\tau = 0.2$ (high cooperativity) for increasing numbers μ of coil states per residue including the limit $\mu \to \infty$. Our results are shown in Figs. 1.6 and 1.7. The crossover behavior for small μ turns into a continuous order-disorder transition as $\mu \to \infty$.

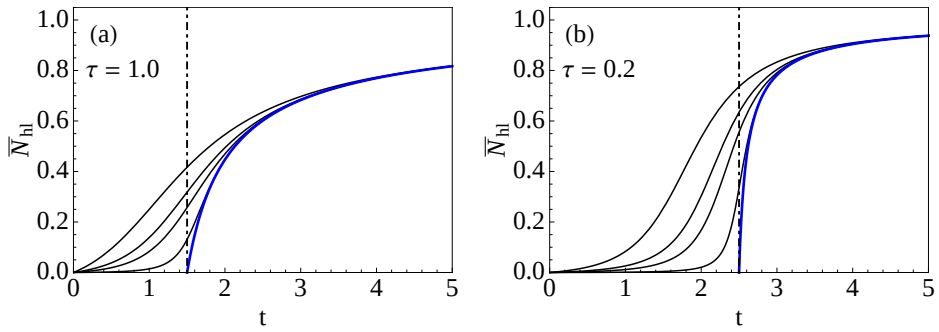

Figure 1.6. (Color online) Helicity $\bar{N}_{\rm hl}$ versus growth parameter t at cooperativity (a) $\tau = 1.0$ and (b) $\tau = 0.2$ for $\mu = 2, 3, 4, 9$ (thin curves from top down) and $\mu = \infty$ (thick curve). The dot-dashed lines marks $t_{\rm c}$.

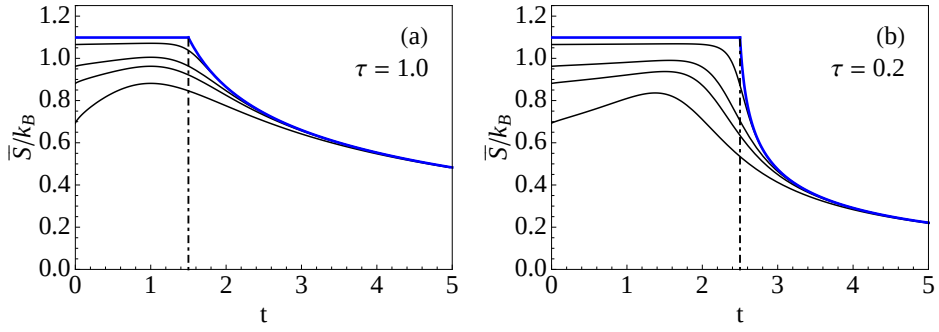

Figure 1.7. (Color online) Entropy $\bar{S}/k_{\rm B}$ versus t at (a) $\tau = 1.0$ and (b) $\tau = 0.2$ for $\mu = 2, 3, 4, 9$ (thin curves from bottom up) and $\mu = \infty$ (thick curve). The dot-dashed line marks $t_{\rm c}$.

With μ increasing, the internal source of disorder in coil segments gains dominance over the entropy of mixing between coil and helix segments. The shoulder in $\bar{S}/k_{\rm B}$ becomes flatter, higher, and sharper. In the limit $\mu \to \infty$ at $t < t_{\rm c}$, the helicity approaches zero identically and the entropy approaches the value $\bar{S}/k_{\rm B} = \ln 3$, independent of τ. Disorder defeats order hands down.

The helix phase at $t > t_{\rm c}$, by contrast, remains a battleground between ordering and disordering tendencies. Both the helicity and the entropy expressions,

$$\bar{N}_{\rm hl} = 1 - \frac{t\tau}{\lambda(2\lambda + 1 - t)}, \tag{1.38}$$

$$\frac{\bar{S}}{k_{\rm B}} = \ln\Big(1+\frac{\tau}{\lambda}\Big) + \frac{t\tau\ln t}{\lambda(2\lambda+1-t)} + \frac{\tau\ln\tau}{\lambda+\tau}\left[\frac{t\tau}{\lambda(2\lambda+1-t)} - 1\right], \tag{1.39}$$

have linear cusps at $t_{\rm c}$ with slopes that diverge in the limit $\tau \to 0$. The leading critical singularities at $t_{\rm c}^{+}$ are

$$\bar{N}_{\rm hl} = \frac{2t_0}{t_{\rm c}}\frac{t-t_{\rm c}}{(t_0-t_{\rm c})^2} + \mathcal{O}\big((t-t_{\rm c})^2\big), \tag{1.40}$$

$$\frac{\bar{S}}{k_{\rm B}} = \ln t_0 - 2\frac{t-t_{\rm c}}{t_0-t_{\rm c}}\left[\frac{\ln t_{\rm c}}{t_0-t_{\rm c}} + \frac{\ln(t_0-t_{\rm c})}{t_{\rm c}}\right] + \mathcal{O}\big((t-t_{\rm c})^2\big). \tag{1.41}$$

With the growth parameter increasing from $t_{\rm c}$ the helicity steeply rises from zero and gradually bends over toward its saturation value whereas the entropy steeply descends from a high value and gradually approaches zero. The plots suggest that cooperativity impedes the onset of ordering, yet assists the quick rise of ordering once it has set in.

1.4.2 Segments of coil and helix

Further insight into how helical ordering grows during the crossover or near the transition point between conformations can be gained from the two quantities (1.17) and (1.18), representing, respectively, the density and average length of segments in one or the other conformation. Coil segments alternate with helix segments. Hence they come in equal numbers. However, their average lengths vary independently with t. Parametric representations can be constructed as before. We use (1.34) and

$$\bar{N}_{\rm seg} = -\frac{\tau}{w_1(1+w_1)}\frac{\partial w_1}{\partial \tau}. \tag{1.42}$$

We first examine the t-dependence of $\bar{N}_{\rm seg}$, $L_{\rm cs}$, and $L_{\rm hs}$ near the first-order transition that takes place at t_0 in the limit $\tau \to 0$. Our results for $\mu = 2, 3$ are shown in Figs. 1.8 and 1.9. We observe that the density of segments is near zero at small t. Here the system is strongly coil-like. The segments grow in numbers with

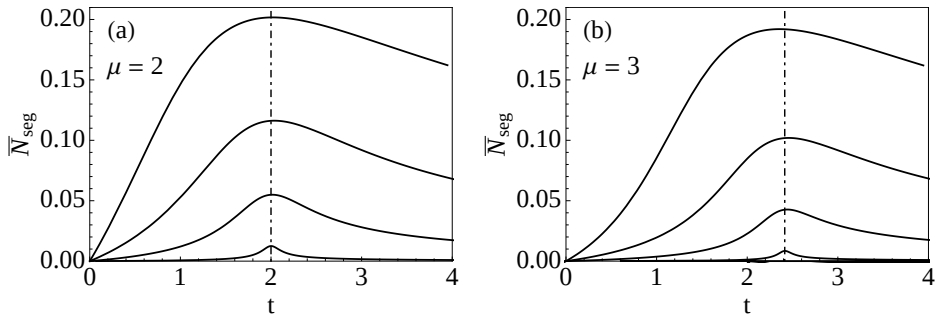

Figure 1.8. Density of segments, $\bar{N}_{\rm seg}$, versus t at $\tau = 1$, 0.25, 0.05, 0.0025 (from top down) for (a) $\mu = 2$ and (b) $\mu = 3$. The dot-dashed line marks t_0.

t increasing. They become most numerous at t_0, where ordering and disordering tendencies compete evenly. The density of segments becomes smaller again as t further increases into the stability regime of the helix conformation. The maximum of $\bar{N}_{\rm seg}$ at t_0 strongly depends on τ. In the limit $\tau \to 0$ we have $\bar{N}_{\rm seg} \equiv 0$, which means that, in a macroscopic system, the number of segments grows more slowly (if at all) than the number of residues.

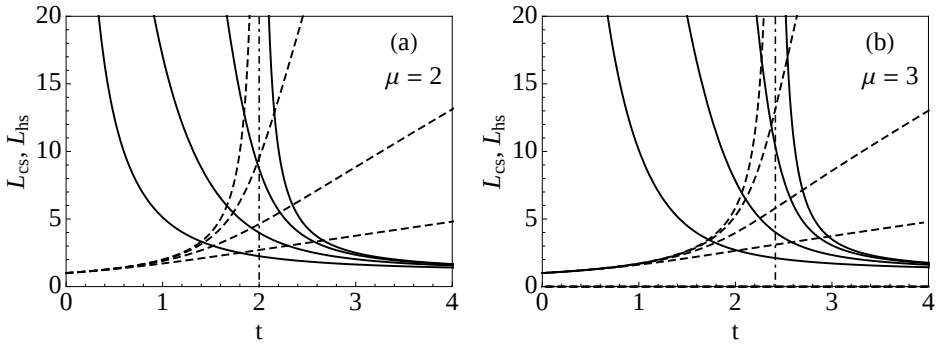

Figure 1.9. Average length of coil segments, $L_{\rm cs}$ (solid lines), and helix segments, $L_{\rm hs}$ (dashed lines), versus t at $\tau = 1$, 0.25, 0.05, 0 (from bottom up) for (a) $\mu = 2$ and (b) $\mu = 3$. The dot-dashed line marks t_0.

Unsurprisingly, $L_{\rm cs}$ decreases and $L_{\rm hs}$ increases monotonically with t. As expected, both variations are enhanced by cooperativity. Most interesting is the limit $\tau \to 0$. The exact expressions for the average lengths of helix segments at $t < t_0$

and coil segments at $t > 0$ read

$$L_{\rm hs} = \frac{t_0}{t_0 - t}, \tag{1.43}$$

$$L_{\rm cs} = \frac{t}{(1+t)(3-t)} \left[\frac{2(\mu+1)}{r_\mu(t-1)} + 2\mu r_\mu(t-1) - (2\mu+1)(t-1) \right], \tag{1.44}$$

respectively, with $r_\mu(w)$ from (1.24). Both expressions diverge $\propto |t - t_0|^{-1}$ as t approaches t_0 from opposite sides and then stay infinite.

These results tell us that the macrostate with zero helicity and saturated entropy at $t < t_0$ still contains helix segments albeit only in numbers that do not add up to a nonzero density but still produce a well-defined average size. They coexist with an equal number of coil segments of macroscopic lengths. Conversely, the macrostate of zero entropy (per residue) and saturated helicity at $t > t_0$ is not a single helical domain. Here helix segments of short average length in numbers that amount to zero density coexist with an equal number of helix segments of macroscopic lengths.

A different picture emerges near the second-order transition at $t = t_{\rm c}$ in the limit $\mu \to \infty$. Results for the density of segments at high and low cooperativity are shown in Fig. 1.10 and results for their average lengths in Fig. 1.11. This includes numerical results for $\mu < \infty$ and analytical results for $\mu = \infty$. The density of segments vanishes identically in the coil phase ($t < t_{\rm c}$) and then rises in a linear cusp to a smooth maximum in the helix phase ($t > t_{\rm c}$) :

$$\bar{N}_{\rm seg} = \frac{\tau}{\lambda + \tau} \left[1 - \frac{t\tau}{\lambda(2\lambda + 1 - t)} \right] = \frac{2t_0}{t_{\rm c}} \frac{t - t_{\rm c}}{t_0 - t_{\rm c}} + \mathcal{O}\big((t - t_{\rm c})^2\big). \tag{1.45}$$

The average length of coil segments in the helix phase,

$$L_{\rm cs} = \frac{t(\lambda + \tau)}{\lambda(2\lambda + 1 - t) - t\tau} = \frac{t_{\rm c}}{2} \frac{t_0 - t_{\rm c}}{t - t_{\rm c}} + \frac{2t_{\rm c}^2 - 9t_{\rm c} + 15}{4(t_0 - t_{\rm c})} + \mathcal{O}(t - t_{\rm c}), \tag{1.46}$$

diverges at $t_{\rm c}^+$ and remains infinite in the coil phase. The average length of helix

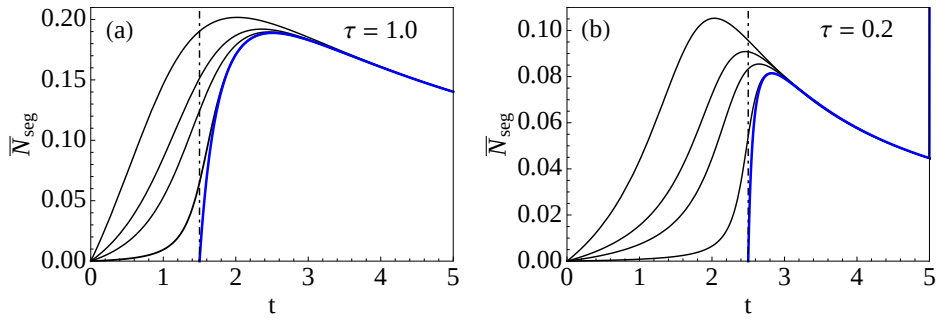

Figure 1.10. (Color online) Density of segments, $\bar{N}_{\rm seg}$, versus t at (a) $\tau = 1.0$ and (b) $\tau = 0.2$ for $\mu = 2, 3, 4, 9$ (thin curves lines left to right) and $\mu = \infty$ (thick curve). The dot-dashed line marks $t_{\rm c}$.

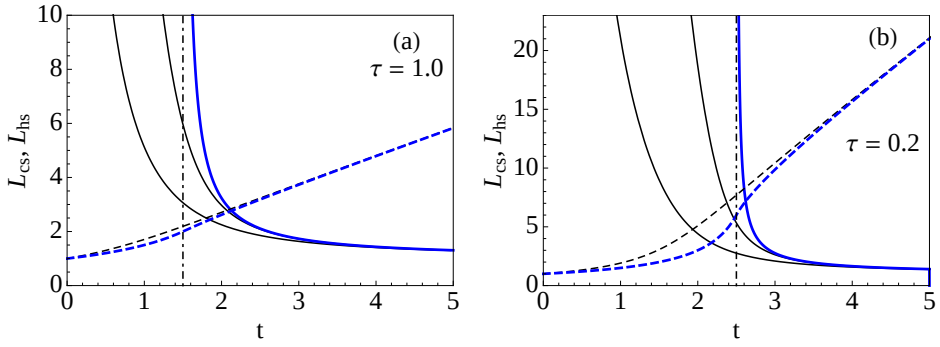

Figure 1.11. (Color online) Average length of coil segments, $L_{\rm cs}$ (solid lines), and helix segments, $L_{\rm hs}$ (dashed lines), for $\mu = 2$ (thin lines) and $\mu = \infty$ (thick lines) versus t at (a) $\tau = 1.0$ and (b) $\tau = 0.2$. The dot-dashed line marks $t_{\rm c}$.

segments, by contrast, remains finite in both phases,

$$L_{\rm hs} = \begin{cases} \dfrac{t_0}{t_0 - t} & : t \le t_{\rm c}, \\ 1 + \dfrac{\lambda}{\tau} & : t \ge t_{\rm c}, \end{cases} \tag{1.47}$$

where again, $t_0 = 3$ for $\mu = \infty$ in (1.45)-(1.47). The graph of $L_{\rm hs}$ is continuous and smooth at $t_{\rm c}$. The singularity is of higher order. Only in the helix phase does the shape of the curve depend on τ.

The most striking feature in the data shown concerns the helix segments. Unlike in the case of the first-order transition, the ordered phase near $t_{\rm c}$ supports a significant density of coil and helix segments of comparable finite size. The

average length of helix segments depends only weakly on μ and only moderately on τ, in strong contrast to the average length of coil segments, which exhibits strong dependences on both parameters.

1.4.3 Heat capacity and latent heat

The heat capacity, $\bar{C} \doteq T(\partial \bar{S}/\partial T)_{\epsilon_n,\epsilon_g}$, illuminates the competition between order and disorder from yet a different angle. From (1.35) we derive

$$\frac{\bar{C}}{k_B} = \frac{2w_1+1}{w_1^2(1+w_1)^2}\left[t\ln t\frac{\partial w_1}{\partial t}+\tau\ln\tau\frac{\partial w_1}{\partial \tau}\right]^2 - \frac{1}{w_1(1+w_1)}\left[t(\ln t)^2\frac{\partial w_1}{\partial t}\right.$$
$$\left.+\tau(\ln\tau)^2\frac{\partial w_1}{\partial \tau}+(t\ln t)^2\frac{\partial^2 w_1}{\partial t^2}+(\tau\ln\tau)^2\frac{\partial^2 w_1}{\partial \tau^2}+2(t\ln t)(\tau\ln\tau)\frac{\partial^2 w_1}{\partial t\partial\tau}\right] \quad (1.48)$$

Figure 1.12 shows the dependence of the heat capacity on the growth parameter for the case $\mu = 2$ at moderate to high cooperativity and for the case $\mu = \infty$ over a wider range of cooperativity.

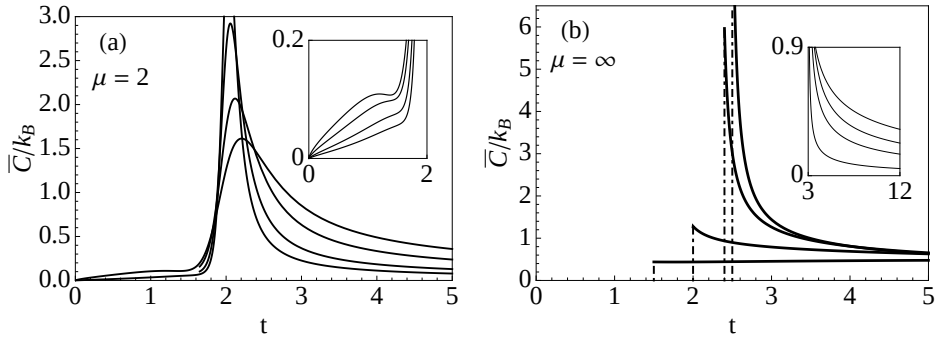

Figure 1.12. Heat capacity $\bar{C}/k_{\rm B}$ versus t for (a) $\mu = 2$ at $\tau = 0.05$, 0.025, 0.01, 0.005 (from top down on the right in the main plot and from bottom up in the inset) and (b) $\mu = \infty$ at $\tau = 1.0$, 0.5, 0.25, 0.2 (from bottom up in main plot) and $\tau = 0.2$, 0.1, 0.05, 0.01 (from top down in the inset). The dot-dashed lines mark t_c for given τ.

At $t < t_0 = 2$ in panel (a) we observe a weak signal that is associated with the entropy caused by alternating coil and helix segments as discussed previously. This contribution fades away at high cooperativity (best seen in the inset) as the density of segments diminishes. The peak at $t \gtrsim t_0$, on the other hand, is

associated with the entropy inside coil segments. With increasing cooperativity, this contribution grows in a more and more narrow range at t_0. Similar structures have been obtained in recent Monte Carlos simulations, albeit upon variation of temperature and in a somewhat different scenario [29].

In the limit $\tau \to 0$ for $\mu < \infty$, where the coil-helix crossover sharpens into a first-order transition at t_0, the heat capacity approaches zero everywhere except at the transition point, where it diverges and produces, via (1.37), a latent heat of magnitude ϵ_g. Conversely, in the limit $\mu \to \infty$ at $\tau > 0$, where the coil-helix crossover turns into a second-order transition at t_c, the heat capacity approaches zero in the coil conformation and remains nonzero in the helix conformation as shown in panel (b). When the transition changes from second to first order when $\tau \to 0$ for $\mu = \infty$, implying $t_c \to t_0 = 3$, the heat capacity throughout the helix conformation approaches zero as illustrated in the inset.

1.5 Conclusion and outlook

We have launched this project mainly for the purpose of interpreting (ongoing and projected) experiments on pHLIP. In this first of three stages of analysis we have constructed a microscopic model for the pH-driven coil-helix conformational change of a long polypeptide adsorbed to a water-lipid interface. We have employed a methodology that facilitates the exact statistical mechanical analysis of our model. The three model parameters t, τ, μ have settings for which the conformation changes either in a crossover, a first-order transition, or a second-order transition.

We have carried out the analysis to the extent needed for a discussion of the sources and agents of order and disorder. Our results include the t-dependence of the helicity (order parameter), the average numbers and the average lengths of helix and coil segments, the entropy, and the heat capacity. The behavior of

these quantities near the continuous or the discontinuous transition has been given special attention. We have plotted all quantities versus t at constant τ and μ for a reason. The experimentally relevant processes for which we use our model will primarily involve variations of the growth parameter t. These variations are caused by changes in pH. The targeted peptides, adsorbed to the water-lipid interface, include side chains that are strongly hydrophobic (e.g. Leu) and side chains that are negatively charged (e.g. Asp, Glu).

A drop in pH leads to the protonation of the negatively charged side chains and, therefore, enhances the overall hydrophobicity. The backbone of a coil segment thus pushed past the lipid headgroups is now more likely to satisfy an H-bond internally than externally. The enthalpic cost for broken internal H-bonds increases. This cost is encoded in t. Any increase in t favors a growth of helix segments at the expense of coil segments. A rise in pH has the opposite effect. The value of t decreases. Coil segments grow at the expense of helix segments.

The cooperativity parameter τ, by contrast, is much less sensitive to a change in pH. In the nucleation process of coil segments from the helix conformation, for example, the internal H-bonds are much more isolated from environmental influences than are those at the border between coil segments and helix segments.

At this point, our project has reached a fork, where natural continuations point in two different directions and address the interests of somewhat different audiences. These continuations, already in the works, are outlined as follows.

1.5.1 Heterogeneous environment and short peptides

In one continuation we begin by considering long polypeptides that are no longer confined to a plane parallel to a flat water-lipid interface. The growth parameter t, which drives the conformational change, then becomes a field $t(x)$ and acquires a profile that depends on the local medium. Here x is a position

coordinate in the direction perpendicular to the plane of the membrane. Such circumstances pose a serious challenge to any existing model and its method of analysis. However, the methodology used here is well positioned in that respect. It has already been proven (in different applications [25, 27, 30]) to be adaptable to heterogeneous environments.

The shape of the parameter field $t(x)$ will be determined by the availability of polar molecules to satisfy external H-bonds along the backbone of the peptide. The dominant factor that shapes the field $t(x)$ will be the density profile $\rho_{\mathrm{w}}(x)$ across the membrane, for which data from experiments [31] and simulations [32] are available. Subdominant factors include electrostatic interactions and fluid-mechanical properties of lipids.

From the analysis of our extended model emerge profiles for the densities of free energy, enthalpy, entropy, and helicity of long polypeptides that traverse the heterogeneous environment (ranging from polar to non-polar) along some path that is subject to conformational constraints [33]. These profiles, in turn, will be interpreted as propensities for the statistical mechanical behavior of short peptides in the same environment.

At this stage of the analysis, additional enthalpic and entropic effects involving the side chains, the semi-fluid bilayer of lipid amphiphiles, and the hydrogen-bonded network of H_2O molecules can be built into the model. The outcome are landscapes of free energy, enthalpy, entropy, and helicity for short peptides of specific composition. The free-energy landscapes in particular then set the stage for (i) a theoretical study of the kinetics of trans-membrane insertion and exit of pHLIP and other membrane peptides and (ii) a direct comparison with experiments currently in progress that investigate the insertion/exit processes of pHLIP via tryptophan fluorescence and the accompanying conformational changes

via circular dichroism spectroscopy.

This first continuation can also benefit from recent studies in the same area of research. Not yet included in our modeling are effects related to torsion and tension, which are bound to be present in the heterogeneous membrane environment. Experimental, computational, and analytic studies of force-extension and torque-twist characteristics and the associated steric constraints [34, 35, 36, 37] will be of great value for that purpose. The kinetic modeling of pHLIP insertion while undergoing a conformational change will find valuable guidance from recent studies that have investigated the fluctuation properties of helical polymers in confined environments including narrow channels [38, 39, 40] and studies that have investigated the Brownian dynamics of polymers in the membrane environment [41].

1.5.2 Extensions of analysis, model, and scope

A second continuation focuses on the statistical mechanics of phase transitions and critical singularities in the context of the microscopic model presented in this work and extensions thereof. It is well known that the presence of a phase transition at nonzero temperature in a system that is, in some sense, one-dimensional requires interactions of long-range to stabilize an ordered phase in the face of strong thermal fluctuations. In our model, which is truly microscopic and analyzed exactly, this stabilizing agent comes in the form of quasiparticles that extend over entire coil segments (hosts) or over parts thereof (hybrids).

In the context of the experiments that motivated this work we have examined conformational changes driven by the control parameter t at fixed τ, μ as reflected in just a few relevant quantities. The phase transitions that occur in the limits $\tau \to 0$ (first-order) or $\mu \to \infty$ (second order) produce different singularities in other quantities of no less interest for the statistical mechanical analysis. Such

quantities of general interest include a mechanical response function, a correlation length, and a correlation function. The further analysis of critical exponents and scaling laws is best presented in a more general framework and along with model extensions that remain inside the reach of our method of exact analysis.

In a final note, we should like to draw the reader's attention to a different set of applications, for which our statistical mechanical model and its extensions are likely to produce significant new insights. These applications investigate the statistical mechanics and the dynamics of DNA melting (thermal denaturation) [13, 14, 42, 43, 44, 45, 46, 47, 48] or the loop formation in RNA [49]. Of particular interest is the loop exponent in the configurational entropy of loop formation [14, 49], which is frequently used as an adjustable parameter. The further development of our project aims for the analytic calculation of loop exponents pertaining to realistic scenarios. Discussions of and debates about crossovers, first-order transitions, and second-order transitions are at the center of most of these studies. The transcription and adaptation of our methodology to this particular physics context is already in progress. The main challenge in the endeavor is the extension of the self-avoiding random walk to three dimensions.

List of References

[1] J.-L. Popot and D. M. Engelman, "Membrane protein folding and oligomerization: the two-stage model," *Biochemistry*, vol. 29, no. 17, pp. 4031–4037, 1990.

[2] S. H. White and W. C. Wimley, "Membrane protein folding and stability: physical principles," *Annual Review of Biophysics and Biomolecular Structure*, vol. 28, no. 1, pp. 319–365, 1999.

[3] D. M. Engelman, Y. Chen, C.-N. Chin, A. R. Curran, A. M. Dixon, A. D. Dupuy, A. S. Lee, U. Lehnert, E. E. Matthews, Y. K. Reshetnyak, *et al.*, "Membrane protein folding: beyond the two stage model," *FEBS Letters*, vol. 555, no. 1, pp. 122–125, 2003.

[4] A. S. Ladokhin and S. H. White, "Interfacial folding and membrane insertion of a designed helical peptide," *Biochemistry*, vol. 43, no. 19, pp. 5782–5791, 2004.

[5] Y. K. Reshetnyak, M. Segala, O. A. Andreev, and D. M. Engelman, "A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers," *Biophysical Journal*, vol. 93, no. 7, pp. 2363–2372, 2007.

[6] D. Weerakkody, A. Moshnikova, M. S. Thakur, V. Moshnikova, J. Daniels, D. M. Engelman, O. A. Andreev, and Y. K. Reshetnyak, "Family of ph (low) insertion peptides for tumor targeting," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5834–5839, 2013.

[7] Y. K. Reshetnyak, O. A. Andreev, M. Segala, V. S. Markin, and D. M. Engelman, "Energetics of peptide (phlip) binding to and folding across a lipid bilayer membrane," *Proceedings of the National Academy of Sciences*, vol. 105, no. 40, pp. 15 340–15 345, 2008.

[8] O. A. Andreev, A. G. Karabadzhak, D. Weerakkody, G. O. Andreev, D. M. Engelman, and Y. K. Reshetnyak, "ph (low) insertion peptide (phlip) inserts across a lipid bilayer as a helix and exits by a different path," *Proceedings of the National Academy of Sciences*, vol. 107, no. 9, pp. 4081–4086, 2010.

[9] A. G. Karabadzhak, D. Weerakkody, D. Wijesinghe, M. S. Thakur, D. M. Engelman, O. A. Andreev, V. S. Markin, and Y. K. Reshetnyak, "Modulation of the phlip transmembrane helix insertion pathway," *Biophysical Journal*, vol. 102, no. 8, pp. 1846–1855, 2012.

[10] D. Poland and H. A. Scheraga, *Theory of helix-coil transitions in biopolymers: statistical mechanical theory of order-disorder transitions in biological macromolecules*. Academic Press, 1970.

[11] B. H. Zimm and J. Bragg, "Theory of the phase transition between helix and random coil in polypeptide chains," *The journal of Chemical Physics*, vol. 31, no. 2, pp. 526–535, 1959.

[12] S. Lifson, "Partition functions of linear-chain molecules," *The Journal of Chemical Physics*, vol. 40, no. 12, pp. 3705–3710, 1964.

[13] D. Poland and H. A. Scheraga, "Phase transitions in one dimension and the helix-coil transition in polyamino acids," *The Journal of Chemical Physics*, vol. 45, no. 5, pp. 1456–1463, 1966.

[14] M. E. Fisher, "Effect of excluded volume on phase transitions in biopolymers," *The Journal of Chemical Physics*, vol. 45, no. 5, pp. 1469–1473, 1966.

[15] F. D. M. Haldane, "Fractional statistics in arbitrary dimensions: A generalization of the pauli principle," *Physical Review Letters*, vol. 67, no. 8, p. 937, 1991.

[16] Y.-S. Wu, "Statistical distribution for generalized ideal gas of fractional-statistics particles," *Physical Review Letters*, vol. 73, no. 7, p. 922, 1994.

[17] S. B. Isakov, "Generalization of statistics for several species of identical particles," *Modern Physics Letters B*, vol. 8, no. 05, pp. 319–327, 1994.

[18] D. Anghel, "The fractional exclusion statistics amended," *EPL (Europhysics Letters)*, vol. 87, no. 6, p. 60009, 2009.

[19] G. G. Potter, G. Müller, and M. Karbach, "Thermodynamics of ideal quantum gas with fractional statistics in d dimensions," *Phys. Rev. E*, vol. 75, p. 061112, 2007.

[20] V. V. Stepanov, G. Müller, and J. Stolze, "Quantum integrability and nonintegrability in the spin-boson model," *Phys. Rev. E*, vol. 77, p. 066202, 2008.

[21] D. Liu, P. Lu, G. Müller, and M. Karbach, "Taxonomy of particles in ising spin chains," *Physical Review E*, vol. 84, no. 2, p. 021136, 2011.

[22] P. Lu, D. Liu, and Müller, "Interlinking motifs and entropy landscapes of statistically interacting particles," *Condensed Matter Physics*, vol. 15, p. 13001, 2012.

[23] D. Liu, J. Vanasse, G. Müller, and M. Karbach, "Generalized pauli principle for particles with distinguishable traits," *Physical Review E*, vol. 85, no. 1, p. 011144, 2012.

[24] N. Gundlach, M. Karbach, D. Liu, and G. Müller, "Jammed disks in a narrow channel: criticality and ordering tendencies," *J. Stat. Mech*, vol. 2013, no. 1, p. P04018, 2013.

[25] C. Moore, D. Liu, B. Ballnus, M. Karbach, and G. Müller, "Disks in narrow channel jammed by gravity and centrifuge: profiles of pressure," *J. Stat. Mech*, vol. 2014, p. P04008, 2014.

[26] B. Bakhti, M. Karbach, P. Maass, M. Mokim, and G. Müller, "Statistically interacting vacancy particles," *Physical Review E*, vol. 89, no. 1, p. 012137, 2014.

[27] A. Meyer, N. Gundlach, M. Karbach, P. Lu, and G. Müller, "Molecular chains under tension and torque: Force-extension and torque-twist characteristics," unpublished.

[28] C. Nowak, V. G. Rostiashvili, and T. A. Vilgis, "Globular structures of a helix-coil copolymer: Self-consistent treatment," *The Journal of Chemical Physics*, vol. 126, no. 3, p. 034902, 2007.

[29] Y. Chen, Q. Zhang, and J. Ding, "A coarse-grained model and associated lattice monte carlo simulation of the coil–helix transition of a homopolypeptide," *The Journal of Chemical Physics*, vol. 120, no. 7, pp. 3467–3474, 2004.

[30] B. Bakhti, M. Maass, M. Karbach, and G. Müller, "Monodisperse rods in external potentials," *Phys. Rev. E*, vol. 92, p. 042112, 2015.

[31] C. Hofsäß, E. Lindahl, and O. Edholm, "Molecular dynamics simulations of phospholipid bilayers with cholesterol," *Biophysical Journal*, vol. 84, no. 4, pp. 2192–2206, 2003.

[32] J. L. MacCallum, W. D. Bennett, and D. P. Tieleman, "Distribution of amino acids in a lipid bilayer from computer simulations," *Biophysical Journal*, vol. 94, no. 9, pp. 3393–3404, 2008.

[33] G. P. Sharma, A. C. Meyer, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, "Free-energy landscapes for peptides in membrane environment," unpublished.

[34] M. N. Tamashiro and P. Pincus, "Helix-coil transition in homopolypeptides under stretching," *Physical Review E*, vol. 63, no. 2, p. 021909, 2001.

[35] F. Hanke, A. Serr, H. J. Kreuzer, and R. R. Netz, "Stretching single polypeptides: The effect of rotational constraints in the backbone," *EPL (Europhysics Letters)*, vol. 92, no. 5, p. 53001, 2010.

[36] B. Chakrabarti and A. J. Levine, "Nonlinear elasticity of an α-helical polypeptide," *Physical Review E*, vol. 71, no. 3, p. 031905, 2005.

[37] B. Chakrabarti and A. J. Levine, "Nonlinear elasticity of an α-helical polypeptide: Monte carlo studies," *Physical Review E*, vol. 74, no. 3, p. 031903, 2006.

[38] A. Lamura, T. Burkhardt, and G. Gompper, "Helical polymer in cylindrical confining geometries," *Physical Review E*, vol. 70, no. 5, p. 051804, 2004.

[39] Y. Yang, T. W. Burkhardt, and G. Gompper, "Free energy and extension of a semiflexible polymer in cylindrical confining geometries," *Physical Review E*, vol. 76, no. 1, p. 011804, 2007.

[40] T. W. Burkhardt, Y. Yang, and G. Gompper, "Fluctuations of a long, semiflexible polymer in a narrow channel," *Physical Review E*, vol. 82, no. 4, p. 041801, 2010.

[41] S. Ramachandran, S. Komura, K. Seki, and G. Gompper, "Dynamics of a polymer chain confined in a membrane," *The European Physical Journal E*, vol. 34, no. 5, p. 46, 2011.

[42] M. Peyrard and A. R. Bishop, "Statistical mechanics of a nonlinear model for dna denaturation," *Physical Review Letters*, vol. 62, no. 23, p. 2755, 1989.

[43] T. Dauxios, M. Peyrard, and A. R. Bishop, "Entropy-driven dna denaturation," *Physical Review E*, vol. 47, no. 684, p. R44, 1993.

[44] N. Theodorakopoulos, T. Dauxois, and M. Peyrard, "Order of the phase transition in models of dna thermal denaturation," *Physical Review Letters*, vol. 85, no. 1, p. 6, 2000.

[45] Y. Kafri, D. Mukamel, and L. Peliti, "Why is the dna denaturation transition first order?" *Physical Review Letters*, vol. 85, no. 23, p. 4988, 2000.

[46] M. S. Causo, B. Coluzzi, and P. Grassberger, "Simple model for the dna denaturation transition," *Physical Review E*, vol. 62, no. 3, p. 3958, 2000.

[47] D. Cule and T. Hwa, "Denaturation of heterogeneous dna," *Physical Review Letters*, vol. 79, no. 12, p. 2375, 1997.

[48] E. Carlon, E. Orlandini, and A. L. Stella, "Roles of stiffness and excluded volume in dna denaturation," *Physical Review Letters*, vol. 88, no. 19, p. 198101, 2002.

[49] T. R. Einert and R. R. Netz, "Theory for rna folding, stretching, and melting including loops and salt," *Biophysical Journal*, vol. 100, no. 11, pp. 2745–2753, 2011.

CHAPTER 2

Free-Energy Landscapes for Peptides in Membrane Environment I: Contributions of the Backbone

2.1 Introduction

This work reports the second stage of a three-stage project aiming at a more profound theoretical and experimental understanding of membrane-associated protein or peptide folding. The first stage, reported in [1], involves the design and solution of a microscopic model for the pH-driven transition of a long polypeptide adsorbed to a water-lipid interface. The methodology used interprets the polypeptide and its (homogeneous, effectively two-dimensional) environment as a system of statistically interacting quasiparticles activated from the (ordered) helix state. These particles represent broken internal H-bonds along the backbone in the (disordered) coil conformation sprawled across the interface in the shape of a self-avoiding random walk.

In one experimentally realized scenario the coil-helix conversion is triggered by a drop in pH. The ensuing protonation of negatively charged side chains enhances the hydrophobicity of the polypeptide and pushes its backbone deeper into the (non-polar) membrane. This environmental change favors the formation of internal hydrogen bonds that stabilize the α-helix conformation. Depending on the settings of the model parameters, our model predicts the conformation to change as a crossover, a first-order transition, or a second-order transition.

Here, in the continuation of this project, we begin by considering long polypeptides that are no longer confined to a flat water-lipid interface but are oriented along some path across the heterogeneous environment comprising the lipid bilayer of a liposome or a biological cell and the surrounding water. The model parameter identified in [1] to drive the conformational change then turns into a

field reflecting the local medium. Such circumstances pose a serious challenge to any existing model and its method of analysis. However, the methodology used here is well positioned in that respect. It has already proven to be adaptable to heterogeneous environments in different applications [2, 3, 4]. Here we extend the model introduced in [1] accordingly.

From this analysis profiles emerge for the densities of free energy, enthalpy, entropy, and helicity of long polypeptides that traverse the membrane along some path consistent with conformational constraints. The profiles reflect enthalpic and entropic consequences of the interactions between the backbone of the polypeptide and the membrane or the surrounding hydrogen-bonded network of H_2O molecules.

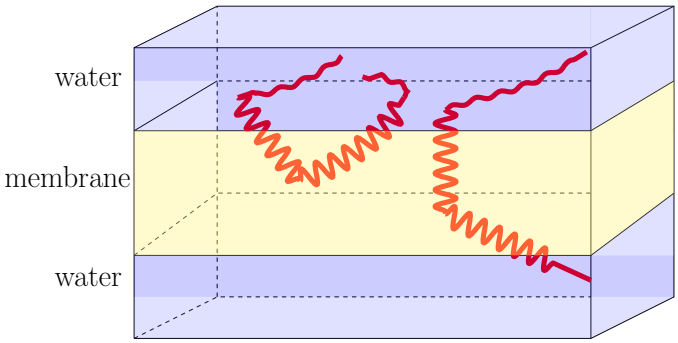

Figure 2.1. (Color online) Short peptide in conformation favored by local membrane environment: two of several configurations analyzed in this work.

In the next step of the analysis these profiles are taken to represent propensities for the statistical mechanical behavior of short peptides in the same environment (see Fig. 2.1). At this stage, enthalpic and entropic effects involving the side chains and the semi-fluid bilayer of lipid amphiphiles can be built into the model. The outcomes are landscapes of free energy, enthalpy, entropy, and helicity for short peptides of given composition. The free-energy landscapes in particular set the stage for a study of the kinetics of trans-membrane insertion and exit of membrane peptides.

This theoretical study is custom-made for variants of the *pH Low Insertion Peptide* (pHLIP) family but not to the exclusion of other peptides with similar attributes. The sequences of three pHLIP variants are depicted in Fig. 2.2 with several features relevant for this work emphasized.

The solubility in water of pHLIP is important. The presence and positioning in the sequence of the charged residues and polar residues is instrumental for this key attribute. The hydrophobic residues provide the affinity for adsorption of pHLIP to the water-lipid interface. The protonatable negative charges at and near the (inserting) C terminus make pHLIP sensitive to the experimentally controllable and reversible change in pH.

The Trp residues are markers for fluorescent spectroscopy, by which the insertion and exit processes are monitored. The three variants with Trp at different positions in the sequence of residues promise to yield clues about the modes of insertion and exit. The conformational changes between coil and helix segments that accompany both insertion and exit are monitored by CD and OCD spectroscopy. The helix-inhibiting Pro residue near the center of the sequence may be instrumental for the mechanics of the insertion process [1].

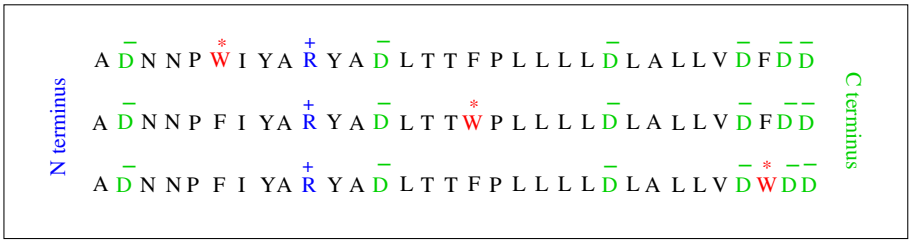

Figure 2.2. (Color online) Sequence of the 32 residues that make up variants W6, W17, W30 (top to bottom) of pHLIP. Highlighted are the residues that have electrically charged side chains and the Trp residue used as a marker in fluorescent spectroscopy.

[1] Experiments that use a variant with the Pro residue (near center) and the Ile residue (near N terminus) switched as a control sample will shed light on this issue.

The following sections are about fields, profiles, and landscapes: fields of environmental parameters (Sec. 2.2), profiles of local properties for long polypeptides (Sec. 2.3), and landscapes of global properties of short peptides (Sec. 2.4).

2.2 Membrane environment

We consider a patch of lipid bilayer with negligible curvature. Heterogeneity is then associated with the spatial coordinate x, perpendicular to the bilayer. As a convention we set $x = 0$ at the center of the bilayer. The outside of the cell or liposome is at positive x and the inside at negative x. Any effects of curvature are higher-order corrections to the results presented in the following.

The membrane environment is characterized by several parameters. In the context of this work the dominant parameter field is the concentration of H_2O molecules. Hydrophobic interactions are prevalent. Subdominant parameter fields involve electrostatic interactions including trans-membrane, surface, and dipole potentials [5]. Further parameter fields are related to properties of lipids, notably the profile of lateral pressure and the entropy reduction along the contact line with the peptide. We begin by examining the effect of the dominant environmental parameter.

2.2.1 Density field of water

We assume that the density field, $\rho_{\rm w}(x)$, of H_2O molecules is symmetric under reflection. It is a dimensionless quantity varying between $\rho_{\rm w}(x) = 1$ sufficiently far from the lipids and $\rho_{\rm w}(x) \simeq 0$ near the center of the bilayer. We use a smoothed-ramp density field as a model representation in our statistical mechanical analysis:

$$\rho_{\rm w}(x) = 1 - \frac{x_s}{x_{\rm a} - x_{\rm b}} \ln \left(\frac{\cosh(x/x_{\rm s}) + \cosh(x_{\rm a}/x_{\rm s})}{\cosh(x/x_{\rm s}) + \cosh(x_{\rm b}/x_{\rm s})} \right) \tag{2.1}$$

It has two adjustable parameters, $x_{\rm b}/x_{\rm a} > 1$ and $x_{\rm s}/x_{\rm a} > 0$. A density field of such shape (Fig. 2.3) is well-established from experiment [6] and computer simulations

[7].

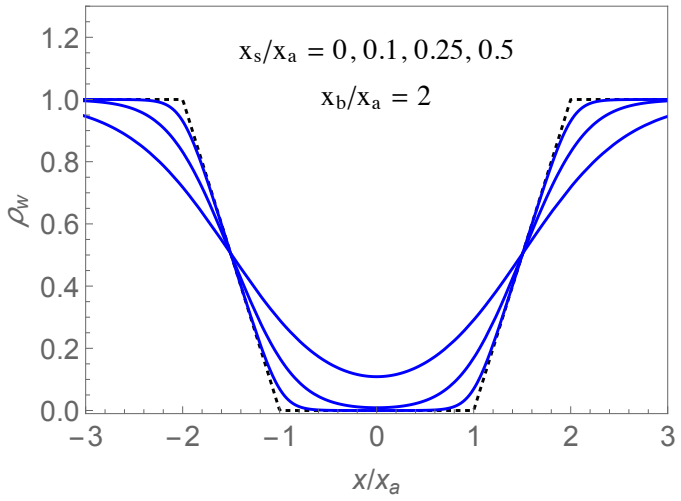

Figure 2.3. Smoothed-ramp model density field of H_2O molecules. The parameters x_a and x_b locate the bottom and the top of the ramp, symmetrically, and x_s is a measure for the softening of the edges.

For our applications we use parameter values as follows: $x_a \simeq 15$Å represents the distance from the center of the bilayer to a point just inside the lipid headgroups, and $x_b \simeq 25$Å the distance from the center to a point outside, where H_2O molecules are in contact with the polar ends of the lipid headgroups; $x_s \simeq 3$Å represents smoothness over an atomic length scale.

A peptide of 35 residues [2] in helix conformation would have a length of roughly 50Å, if we assume that each helical link adds an element $l_h \simeq 1.5$Å of length in the direction of the axis [8]. The length of the peptide in the coil conformation is a fluctuating quantity. The actual size of each link is $l_a \simeq 4$Å. If the peptide is fully extended it reaches the contour length of ~ 140Å, almost three times its length in helix conformation.

In a homogeneous medium the coil conformation can be modelled by a self-avoiding random walk. The average end-to-end distance predicted by an unre-

[2]The most widely used pHLIP variant has 35 residues.

stricted random walk is $\sim$ 24Å, about half of its value in helix conformation. Geometrical and dynamical constraints [9] make the average end-to-end distance somewhat longer. In the membrane environment the average end-to-end distance between N terminus and C terminus is much harder to estimate and best handled case by case.

2.2.2 Free-energy

The term *free-energy landscape* as introduced in the title and then invoked throughout this work is in need of some explanation. The system under consideration includes a peptide in an environment consisting of a lipid bilayer surrounded by water. The thermodynamic potential in use is the Gibbs free energy G. For a homogeneous system it can be expressed in the form,

$$G(T,p) = \underbrace{U + pV}_{H} - TS, \tag{2.2}$$

where U is the internal energy, H the enthalpy, S the entropy, and V the volume. The control variables are the pressure p and the temperature T.

Our system is homogeneous in some but not all respects. We only consider situations at uniform T, typically room temperature. The pressure is uniform in the aqueous environment and its normal component, p_{N}, also across the bilayer. However, the lateral component, p_{L}, has a characteristic profile that averages out to the value of the normal pressure.

For the purpose of this study we only consider quasistatic processes of a restricted type in which both T, and p_{N} remain constant, at ambient pressure and room temperature, respectively. One natural energy scale, therefore, uses $k_{\mathrm{B}}T_{\mathrm{rm}} \simeq 4.0 \times 10^{-21}\mathrm{J} \simeq 0.58\mathrm{kcal/mol}$ (at $T_{\mathrm{rm}} = 293\mathrm{K}$) as its unit. These processes involve the translocation of the peptide, accompanied by changes in conformation of the peptide and in its interactions with water and lipids.

Each such process can be described as a path in some parameter space. The variation of G along any such path is then a section of a free-energy landscape. Path segments where $\Delta G < 0$ are favorable and path segments with $\Delta G > 0$ unfavorable regarding spontaneous occurrence. Paths that are all 'downhill' are likely to have fast realizations in experiments and paths that have significant barriers between initial and final points may be realized only on much slower time scales.

In much of this study we categorize all changes of G as either enthalpic or entropic in nature. We write,

$$\Delta G = \Delta H - T\Delta S, \tag{2.3}$$

and refer to changes with $\Delta H < 0$ as producing an enthalpic gain and changes with $\Delta S > 0$ as producing an entropic gain. In both cases a *gain* is thus associated with a negative contribution to the free energy and *loss* with a positive contribution.

Entropic losses, $\Delta S < 0$, under examination in this work include the following kinds: (i) a complete or partial conformational change of the peptide from (disordered) coil to (ordered) helix; (ii) the immobilization of H_2O molecules via the formation of H-bonds with polar contacts on the backbone or the side chains of the peptide or with polar contacts on lipid head-groups; (iii) formation of an ordered contact line between lipid head-groups and the peptide in adsorbed or trans-membrane state.

Enthalpic losses related to a change $\Delta U > 0$ are all associated with molecular interactions. The two main sources of this type in the context of our study involve (i) the breaking of H-bonds and (ii) the translocation of charges or polar contacts from a polar environment (water) into a non-polar environment (membrane). The H-bonds in question include internal ones between backbone contacts, and H-bonds between the peptide (backbone or side chains), H_2O molecules, and lipid head-groups, all of which have polar contacts. A different type of enthalpic loss,

$(p_{\mathrm{L}}^{\mathrm{h}} - p_{\mathrm{L}}^{\mathrm{l}})\Delta V > 0$, comes into play when a peptide segment (e.g. a residue) of volume ΔV translocates from a position of low lateral pressure, $p_{\mathrm{L}}^{\mathrm{l}}$, to position of high lateral pressure, $p_{\mathrm{L}}^{\mathrm{h}}$.

It is quite challenging, in general, to estimate all these contributions with some accuracy. Existing estimates found in the literature vary widely and often contradict each other, in part due to differences in underlying assumptions. In some instances, several contributions are lumped together such as in tabulated data for hydrophobic transfer free energies.

2.2.3 Enthalpic cost of H-bonds

In the α-helix conformation the backbone of each residue is involved in two H-bonds. The CO group of residue n is acceptor to the NH group of residue $n+4$ acting as donor (Fig. 2.4). The helix conformation thus involves one internal H-bond per residue. In the conversion of a helix segment into a coil segment, a number of internal H-bonds must be broken, for which there is an enthalpic cost.

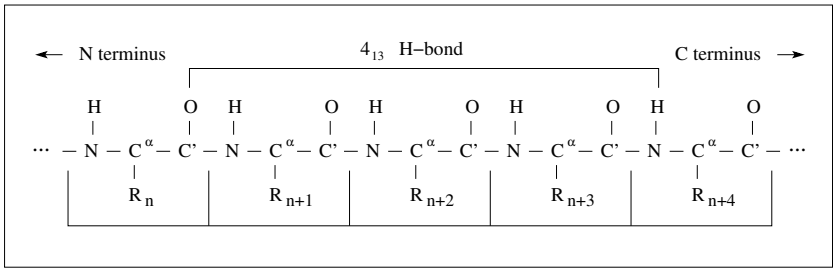

Figure 2.4. Schematic of internal 4_{13} H-bond (bridging 4 residues and 13 chemical bonds along the backbone) of an α-helix.

Deep inside the lipids the full price is due, $\lesssim 9k_{\mathrm{B}}T$ per H-bond, which is considerable in units of ambient thermal fluctuations. Closer to the lipid-water interface, where $\rho_{\mathrm{w}}(x)$ is significant, some broken internal H-bonds may be replaced by H-bonds with H_2O molecules in the role of donors or acceptors as needed.

When the peptide is in coil conformation and positioned in the polar environ-

ment of the lipid headgroups and the adjacent water, there are opportunities for all internal H-bonds to be replaced external ones. This increases the maximum number of H-bonds per residue from one to two. Each residue is again donor and acceptor, but now via bonds not shared with other residues.

Whether the enthalpic bottom line in this case is a gain or loss depends on how the exposed backbone of the polypeptide affects the internal H-bonds of liquid water. It is not obvious which is the case. In ice there are two intact H-bonds per H_2O molecule. Each molecule shares four bonds, two in the role of donor and two as acceptor. In liquid water about 12% of H-bonds between H_2O molecules are broken [8] [3]. The intact H-bonds form a dynamic network with H_2O molecules on average sharing less than four bonds.

If one internal H-bond on the backbone of the peptide is replaced by two external H-bonds involving H_2O molecules from surrounding liquid water, does the average number of H-bonds increase, decrease, or stay the same? The answer depends on the average number of H-bonds in the H_2O network that are broken when two external H-bonds (between one residue and two H_2O molecules) are formed. The enthalpic cost of the helix-coil transformation is zero if that number is equal to one. Any higher number results in an enthalpic loss and any lower number in an enthalpic gain.

If the fraction of unsatisfied H-bonds between H_2O molecules is higher at the lipid-water interface than inside the bulk, then the exposed CO groups and NH groups along a coil segment of the backbone are more likely to encounter partners for external H-bonds in adjacent H_2O molecules. This reduces the need for breaking H-bonds between H_2O molecules when external H-bonds are formed along the backbone of the peptide. The formation of external H-bonds in water is

[3]This estimate compares the latent heat of sublimation, $L_{\rm sg} = 680$cal/g to the latent heat of melting, $L_{sl} = 80$cal/g. If both latent heats are dominated by the breaking of H-bonds, then the fraction of unsatisfied H-bonds in the liquid phase beomes $80/680 \simeq 0.12$.

then more likely to result in an enthalpic gain than loss.

In the context of the methodology developed in [1], the enthalpic contribution to the conformational tendency in the heterogeneous membrane environment is accounted for by turning the activation energies ϵ of coil links from the helix reference state into fields [4]. We use the general ansatz,

$$\epsilon(x) = \epsilon_0\big[1 - \alpha\rho_{\rm w}(x)\big], \tag{2.4}$$

for their dependence on the density field of water, $\rho_{\rm w}(x)$, which represents the dominant environmental influence.

Near the center of the lipid bilayer we have $\rho_{\rm w}(x) \ll 1$, which maximizes $\epsilon(x)$ to roughly the strength of an internal H-bond. At positions closer to the lipid-water interface, $\epsilon(x)$ decreases as $\rho_{\rm w}(x)$ increases. This change is due to the growing probability that internal H-bonds can be replaced by external ones. The parameter α determines whether in the aqueous environment, with $\rho_{\rm w}(x) \simeq 1$, we have an enthalpic gain ($\alpha > 1$) or an enthalpic loss ($\alpha < 1$) [5].

2.2.4 Entropic cost of H-bonds

The enthalpic cost reduction associated with external H-bonds in the polar environment of liquid water comes with an entropic price that has yet to be included in the accounting. Every H-bond formed between an exposed backbone CO or NH group with an H_2O molecule immobilizes that water molecule and thus lowers its contribution to the entropy.

It is hard to estimate this entropy reduction from first principles but, in all likelihood, it is somewhat larger in magnitude than the entropic gain per residue

[4] We use the name ''link' for the unit of peptide backbone between successive peptide bonds as used in the mathematical modelling of Ref. [1].

[5] For the strength of hydrogen bonds we are using, for the time being, one universal value, $\epsilon_{\rm Hb}/k_{\rm B}T \simeq 9$, irrespective of the context. It represents the commonly cited 5kcal/mol. This is an oversimplification. The strength of H-bonds varies and depends on geometrical constraints and the strength of the polar contacts.

produced when a segment of (ordered) helix transforms into a segment of (disordered) coil. In [1] we have calculated that entropy gain per residue to range between $\ln 2 \simeq 0.7$ and $\ln 3 \simeq 1.1$ in units of $k_{\rm B}$. Residues in coil conformation are clearly more constrained than H_2O molecules in the dynamic network of H-bonds that make up liquid water. For what follows we have set

$$\frac{\Delta S_{\rm H}}{k_{\rm B}} \simeq 1.5 \tag{2.5}$$

which we estimate to be an upper limit.

The main message for Sec. 2.4 below is that the entropic contribution to the free energy of the peptide backbone in coil conformation is likely to be positive, i.e. unfavorable, when the peptide is contact with water. Here we have one factor favoring insertion. It will be weighed against other factors yet to be identified and examined [6].

2.2.5 Model for peptide conformation

Previously [1], we introduced and solved a microscopic statistical mechanical model for the conformational transformation between coil and helix of a long polypeptide positioned in a plane parallel to the interface between a polar and a non-polar fluid medium such as realized by water and a lipid bilayer. Depending on the settings and variations of its parameters, the model predicts a crossover or a transition of first or second order between coil and helix.

All microstates of the peptide are characterized in our model by combinations of 2μ species of statistically interacting nested particles: hosts ($m = 1$) nucleate coil segments whereas hybrids ($m = 2, \ldots, \mu$) and tags ($m = \mu + 1, \ldots, 2\mu$) grow such segments in two perpendicular directions of a self-avoiding random walk as illustrated in Fig. 2.5.

[6] The value (2.5) is provisional, probably representing an upper limit. There are reasons to argue that the number of immobilized H_2O molecules is smaller than two per residue. That would reduce the value of parameter α.

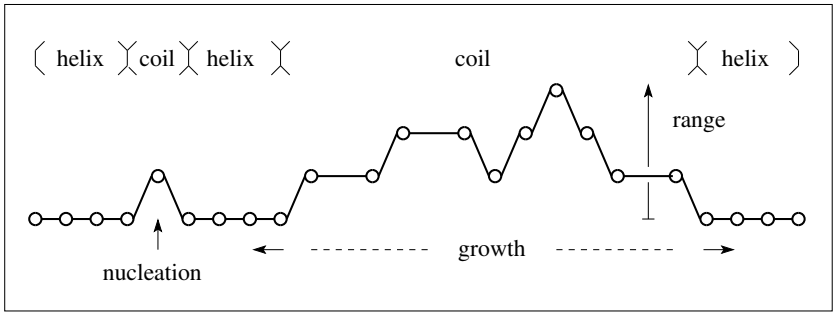

Figure 2.5. Coil conformation of polypeptide modeled as a self-avoiding random walk generated by statistically interacting nested particles activated from the helix reference state via nucleation (control parameter τ) and growth (control parameter t) with range limited by control parameter μ.

The activation energies ϵ_m, $m = 1, \ldots, 2\mu$ can be tailored to meet the physical requirements at hand. In the plane of the water-lipid interface all activation energies are spatially uniform. Nucleating a coil segment requires that several internal H-bonds along the backbone of the α-helix are broken simultaneously whereas the growth of a coil segment already nucleated proceeds by the sequential breaking of single H-bonds.

The model solved in [1], therefore, assigns different activation energies for the control of nucleation than for the control of growth, namely $\epsilon_1 \doteq \epsilon_{\rm n}$ for hosts, $\epsilon_2 = \cdots \epsilon_\mu \doteq 2\epsilon_{\rm g}$ for hybrids, and $\epsilon_{\mu+1} = \cdots \epsilon_{2\mu} \doteq \epsilon_{\rm g}$ for tags. These activation energies in units of the thermal energy $k_{\rm B}T$ are usefully expressed by the nucleation parameter τ (also named cooperativity) and the growth parameter t:

$$\tau = e^{\beta(\epsilon_{\rm g} - \epsilon_{\rm n})} \quad : \ 0 \leq \tau \leq 1, \tag{2.6a}$$

$$t = e^{\beta\epsilon_{\rm g}} \quad : \ 0 \leq t < \infty. \tag{2.6b}$$

In addition to these two continuous control parameters, the discrete model parameter μ controls the range of the random walk away from the axis of the local helix segments. All model features used in the following are reviewed in Appendix B.

2.2.6 Model parameter field

Of the three model parameters the growth parameter t is the one most sensitive to the environment by far. We begin our model adaptation to the membrane environment by keeping the cooperativity parameter τ and the number μ of coils states per residue uniform, while we turn t into a field.

For this purpose we now use the ansatz (2.4) for the construction of two fields of scaled activation energy, one for hosts, the other for hybrids and tags [7]. The former ($m = 1$) we leave undetermined and the latter ($m = 2, \ldots, 2\mu$) we link to the density field of water:

$$K_1(x) \doteq \frac{\epsilon_1(x)}{k_{\mathrm{B}}T}, \tag{2.7}$$

$$K_{\mathrm{t}}(x) \doteq \frac{\epsilon_{\mathrm{Hb}}}{k_{\mathrm{B}}T}\big[1 - \alpha_{\mathrm{H}}\rho_{\mathrm{w}}(x)\big], \tag{2.8}$$

where $\epsilon_{\mathrm{Hb}}/k_{\mathrm{B}}T$ represents the scaled energy of an H-bond and $\alpha_{\mathrm{H}} \simeq 1$ is the enthalpy parameter introduced previously, now assumed to be equal for hybrids and tags.

The growth parameter field,

$$t(x) = e^{K_{\mathrm{t}}(x)} \qquad : \ 0 \leq t(x) < \infty, \tag{2.9}$$

is environmentally sensitive via the shape of $\rho_{\mathrm{w}}(x)$ and the value of α_{H}. The cooperativity,

$$\tau = e^{K_{\mathrm{t}}(x) - K_1(x)} \qquad : \ 0 \leq \tau \leq 1, \tag{2.10}$$

is kept as a continuous model parameter controlling the nucleation of coil segments. The function $K_1(x)$ is then determined via (2.10).

Our reasoning for this choice of modelling is that cooperativity, which controls the nucleation of coil segments, is a process initiated by thermal fluctuations within the backbone of an intact segment of α-helix. Multiple H-bonds must be

[7] We keep setting $\epsilon_{\mathrm{h}} = 2\epsilon_{\mathrm{t}}$ for the activation energies of hybrids ad tags, respectively as in Ref. [1] for the same reason.

broken simultaneously. They are all protected from immediate contact with the environment. Nucleation is affected indirectly by an environmental change from non-polar to polar. In the non-polar environment the nucleation energy barrier is followed by a high plateau and in the polar environment by a low plateau. The difference is accounted for by the growth parameter field.

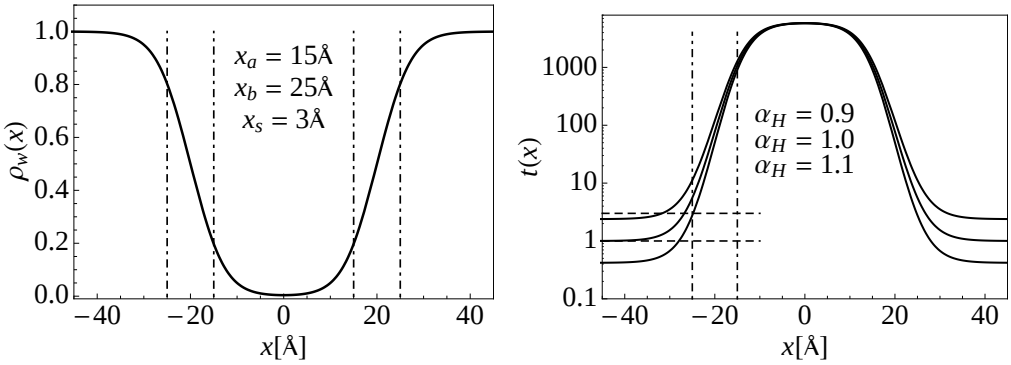

Figure 2.6. Model density field of water $\rho_{\rm w}(x)$ [left] and growth parameter field $t(x)$ [right], with specifications as indicated. The dot-dashed lines represent the endpoints of the ramp and the dashed lines the range $1 \leq t \leq 3$ in which a coil-helix phase transition may occur.

The panel on the left in Fig. 2.6 shows the model density field of water $\rho_{\rm w}(x)$ used henceforth. It has the smoothed-ramp profile (2.1). The growth parameter field $t(x)$ inferred from the predominant environmental field $\rho_{\rm w}(x)$ via (2.8) and (2.9) is shown in the panel on the right. When the polypeptide is in coil conformation while adsorbed to the water-lipid interface its position is near the outer dot-dashed line.

For the coil conformation to be stable, the growth parameter must be below the upper dashed line (at $t = 3$) at the very least. We see that is not the case unless $\alpha_{\rm H} > 1$, i.e. unless the breaking internal H-bonds along the backbone of the polypeptide and replacing them by external H-bonds with available H_2O molecules is an enthalpic gain.

Now we have the extended model system ready for applications to the hetero-

geneous membrane environment. We have converted t into the field (2.9) and keep the control parameters τ and μ as in [1]. We have already stated reasons not to turn the nucleation parameter τ into a field. Regarding discrete model parameter μ [8] we will primarily consider the cases $\mu = 2$ and $\mu = \infty$, for which analytic solutions are available to our analysis in extension to the solutions presented in [1]. The two values span a range that is correlated with a range of entropy generated inside coil segments of given length.

2.3 Profiles

The analysis as carried out in the following yields profiles for some relevant local attributes of peptides in this heterogeneous membrane environment: densities of free energy $\bar{G}$, enthalpy $\bar{H}$, entropy $\bar{S}$, and helicity $\bar{N}_{\rm hl}$. These profiles are still attributes of long, generic polypeptides. What is taken into account at this stage are the internal H-bonds along the backbone of the polypeptide and external H-bonds with water or lipid headgroups depending on their availability. Also taken into account is the entropic effect the exposed backbone in coil conformation has on the surrounding water or lipid molecules.

The polypeptide is oriented perpendicular to the membrane and the length per residue is taken to be independent of the conformation. We can then use the expressions for $\bar{N}_{\rm hl}(t,\tau)$, $\bar{S}(t,\tau)$, $\bar{G}(t,\tau)$, and $\bar{H}(t,\tau)$ from Appendix B with $t = t(x)$ from (2.8), (2.9) and any choice of τ. The growth parameter field $t(x)$ determines all profiles via local relations. Profiles for $\mu = 2$ are shown Fig. 2.7 and profiles for $\mu = \infty$ in Fig. 2.8.

Well inside the lipid bilayer, at $x \simeq 0$, the helix conformation is firmly es-

[8] It is fair to argue that the coil becomes more constrained in the lipid environment than in the aqueous environment, which suggest that we should use a larger value of μ in the latter than in the former. To implement the restricted space available to the polypeptide coil in this way is difficult from a technical point of view. It would entail approximations of little control and is likely to obscure other environmental effects. We, therefore, opt to account for the space restrictions imposed by the lipids in different ways.

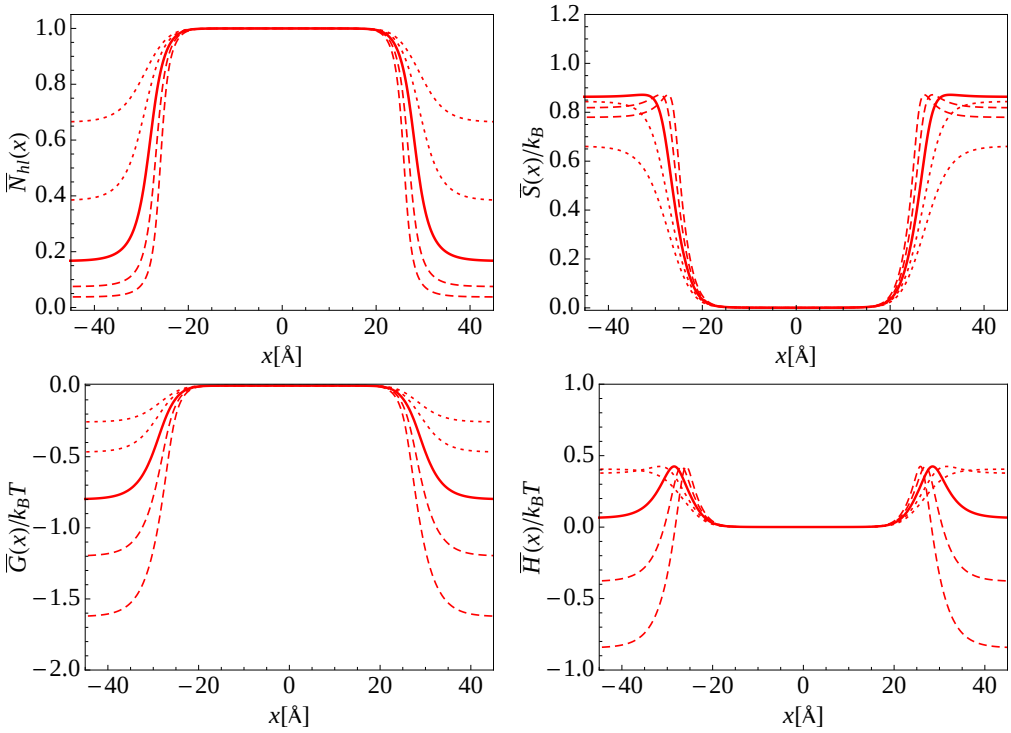

Figure 2.7. Profiles of helicity, entropy, free energy, and enthalpy for a long polypeptide. The model parameter values are $\mu = 2$ and $\tau = 0.5$. The growth parameter field $t(x)$ uses (2.8)-(2.9) with $\rho_{\rm w}(x)$ from Fig. 2.6. The solid, dashed, and dotted curves pertain to $\alpha_{\rm H} = 1$, $\alpha_{\rm H} = 1.05, 1.1$, and $\alpha_{\rm H} = 0.95, 0.9$, respectively.

tablished. All internal H-bonds are intact. There is no configurational disorder. Therefore, the order parameter (helicity) is very close to saturation whereas the densities of enthalpy and entropy are near zero. In consequence, the free-energy density of the polypeptide rises only imperceptibly above its (zero) reference value as well. As we move the position x away from the center of the bilayer, out of the membrane environment into the aqueous environment, the helicity decreases and the entropy increases, the former reflecting a drop in (helical) order and the latter a rise in (coil-like) disorder, both associated with the same conformational change.

In [1] we have identified one source of order and two sources of disorder involving the secondary structure of the polypeptide alone. Order increases with the growth of segments of helix conformation. Disorder is contained in the positions

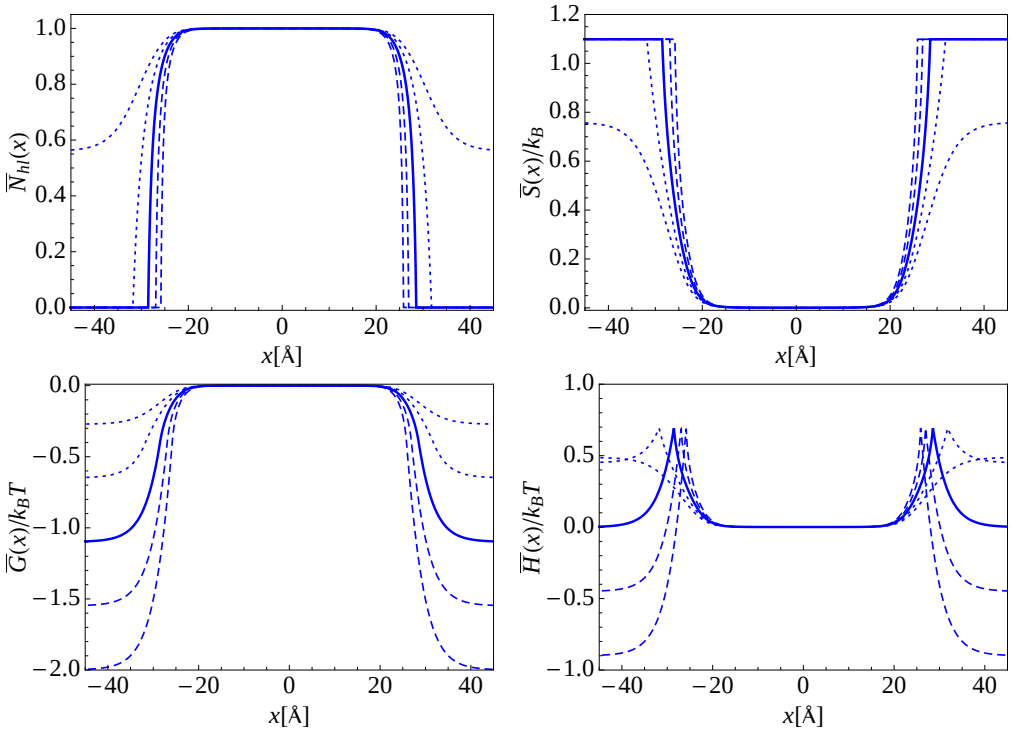

Figure 2.8. Profiles of helicity, entropy, free energy, and enthalpy for a long polypeptide. The model parameter values are $\mu = \infty$ and $\tau = 0.5$. The growth parameter field $t(x)$ uses (2.8)-(2.9) with $\rho_{\rm w}(x)$ from Fig. 2.6. The solid, dashed, and dotted curves pertain to $\alpha_{\rm H} = 1$, $\alpha_{\rm H} = 1.05, 1.1$, and $\alpha_{\rm H} = 0.95, 0.9$, respectively.

of boundaries between coil segments and helix segments as well as inside each coil segment.

The enthalpy density exhibits a sharp maximum at locations near the lipid-water interface. At the peak position the thermal fluctuations are just strong enough to break the internal H-bonds but the environment is not yet sufficiently polar to be able to replace all of them by external H-bonds. Outside the interface the enthalpy profile levels off in a high or low plateau depending on the value of the physical parameter $\alpha_{\rm H}$. That parameter also affects the drop in helicity and the rise in entropy. Enthalpic loss ($\alpha_{\rm H} < 1$) favors order and suppresses disorder in the aqueous environment.

There are some qualitiative and some quantitative differences between the

curves for $\mu = 2$ shown in Fig. 2.7 and those for $\mu = \infty$ shown in Fig. 2.8. The case $\mu = \infty$ produces pure coil with maximum entropy in the aqueous environment. Pure coil means zero helicity. The entropy that can be generated in coil segments by the case $\mu = \infty$ is significantly higher than that generated by the case $\mu = 2$.

The enthalpic spikes near the interface are more pronounced in the case $\mu = \infty$. This difference is attributable to an entropic effect. The breaking of an H-bond at significant enthalpic cost is more likely to happen if the entropy produced is large ($\mu = \infty$) than if it is small ($\mu = 2$).

The free-energy profiles in Figs. 2.7 and 2.8 tell us that the incentives for the insertion of peptides must come from a source other than what has already been taken into account. The free-energy density $\bar{G}(x)$ is significantly higher in the membrane environment than in the aqueous environment. The enthalpic balance strongly depends on the parameter $\alpha_{\rm H}$ but it is not decisive. The entropic contribution strongly favors coil over helix, which, in this context, means water over lipids. At this stage of the analysis we consider one source that modifies the free-energy density profile toward favoring insertion. Further sources, associated with side chains and their interaction with lipids.

The replacement of backbone internal H-bonds with external H-bonds that immobilize H_2O molecules from the aqueous environment, while providing an enthalpic discount as already accounted for, comes at a significant entropic cost. This effect can be taken into account via an amended free-energy density constructed as follows:

$$\frac{\bar{G}(x)_{\rm H}}{k_{\rm B}T} = \frac{\bar{G}(x)}{k_{\rm B}T} + 2\frac{\Delta S_{\rm H}}{k_{\rm B}}\Big[1 - \bar{N}_{\rm hl}(x)\Big], \tag{2.11}$$

with $\Delta S_{\rm H}$ from (2.5). The factor in square brackets represents the fraction of backbone segments in coil conformation with each segment offering docks for two H_2O molecules.

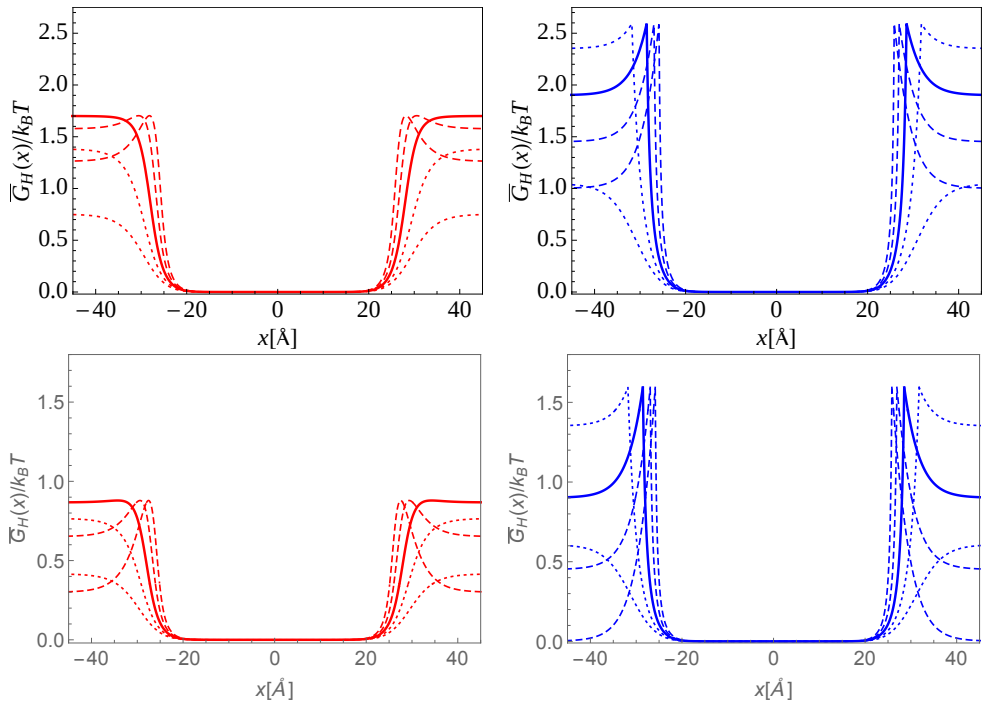

Figure 2.9. Profiles of free energy for a long polypeptide including the entropic contribution from external H-bonds. The model parameter values are $\tau = 0.5$, $\Delta S_{\rm H}/k_{\rm B} = 1.5$ [top] and $S_{\rm H}/k_{\rm B} = 1.0$ [bottom] from(2.6), and $\mu = 2$ (left), $\mu = \infty$ (right). The growth parameter field $t(x)$ uses (2.8)-(2.9) with $\rho_{\rm w}(x)$ from Fig. 2.6. The solid, dashed, and dotted curves pertain to $\alpha_{\rm H} = 1$, $\alpha_{\rm H} = 1.05, 1.1$, and $\alpha_{\rm H} = 0.95, 0.9$, respectively.

The profiles for the amended free-energy densities are shown in Fig. 2.9 for two values of $\Delta S_{\rm H}/k_{\rm B}$. We recall from earlier that the (higher) value (2.7) has been estimated to be an upper limit. This amendment does indeed produce an incentive for insertion. It shifts the global minimum of the free-energy density to the middle of the lipid bilayer. Most interestingly, the modified profiles exhibit, except for low values of $\alpha_{\rm H}$, an energy barrier at the interface between membrane and water.

Note also that for $\alpha > 1$ the global minimum switches position from lipid environment to water environment as the $\Delta S_{\rm H}/k_{\rm B}$ is lowered below 1.0. The significance of these features for stable and metastable states of short peptides will

be further investigated below.

2.4 Landscapes

At this next stage of the statistical mechanical analysis we interpret the profiles calculated in Sec. 2.3 as propensities of residues in short peptides such as pHLIP. The density profiles represent average local attributes of long, generic polypeptides in and around the membrane environment. Here we use them as one factor affecting the behavior of residues of short peptides in the same environment, specifically their conformational preference (coil or helix). Other factors depend on attributes of the specific side chains in the short peptide and further (enthalpic and entropic) effects of peptide-lipid interactions.

The analysis to be carried out produces landscapes of global properties of short peptides: free energy, enthalpy, entropy, helicity again and others. From the shapes of these landscapes with all major contributions included, in particular from the global and local minima of the free-energy landscape, we glean information about the preferred conformation, orientation, and position of the short peptides in the heterogeneous membrane environment. In this Sec. 2.4 we discuss ways to construct landscapes for short peptides in general and, for one scenario, what the contributions from their backbones are.

2.4.1 Scenario #1

Consider a peptide of N_{R} residues oriented perpendicular to the plane of the bilayer. The residues are positioned along that normal as shown schematically in Fig. 2.10. The peptide position x_0 is taken to be the position of the center residue if N_{R} is odd and the midpoint of the center link if N_{R} is even. The relative positions of all residues are then generated recursively from x_0 out toward both

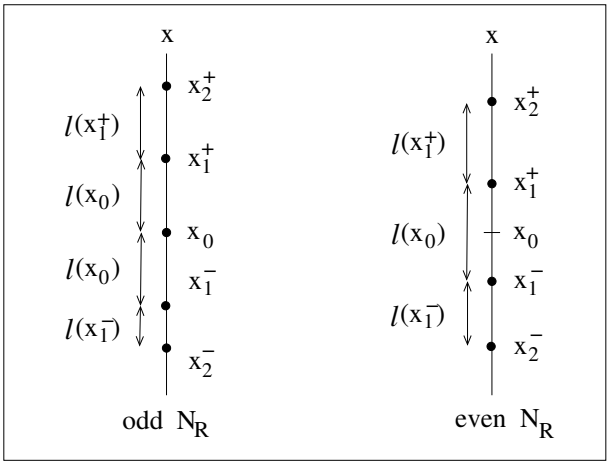

Figure 2.10. Specifications for the positions of residues along x-axis perpendicular to the plane of the bilayer.

ends as follows:

$$x_0^{(\pm)} = x_0, \quad x_n^{(\pm)} = x_{n-1}^{(\pm)} \pm l\left(x_{n-1}^{(\pm)}\right), \tag{2.12}$$

for $n = 1, \ldots, \frac{1}{2}(N_{\mathrm{R}} - 1)$ if N_{R} is odd, and

$$\begin{aligned} x_1^{(\pm)} &= x_0 \pm \tfrac{1}{2} l(x_0), \\ x_n^{(\pm)} &= x_{n-1}^{(\pm)} \pm l\left(x_{n-1}^{(\pm)} \pm \tfrac{1}{2} l\left(x_{n-1}^{(\pm)}\right)\right), \end{aligned} \tag{2.13}$$

for $n = 2, \ldots, \frac{1}{2}N_{\mathrm{R}}$ if N_{R} is even. The function $l(x)$ is a part of the modeling that varies with context (see Sec. 2.4.3 below). The expression for free energy per residue and for helicity can then be calculated from the profile functions evaluated at specific positions via

$$\frac{\bar{G}_{\mathrm{H}}^{(\mathrm{p})}(x_0)}{k_{\mathrm{B}}T} = \frac{1}{N_{\mathrm{R}}}\left[\frac{\bar{G}_{\mathrm{H}}(x_0)}{k_{\mathrm{B}}T} + \sum_{n=1}^{\frac{1}{2}(N_{\mathrm{R}}-1)} \left\{ \frac{\bar{G}_{\mathrm{H}}\left(x_n^{(+)}\right)}{k_{\mathrm{B}}T} + \frac{\bar{G}_{\mathrm{H}}\left(x_n^{(-)}\right)}{k_{\mathrm{B}}T} \right\}\right], \tag{2.14}$$

$$\bar{N}_{\mathrm{hl}}^{(\mathrm{p})}(x_0) = \frac{1}{N_{\mathrm{R}}}\Big[\bar{N}_{\mathrm{hl}}(x_0) + \sum_{n=1}^{\frac{1}{2}(N_{\mathrm{R}}-1)} \left\{\bar{N}_{\mathrm{hl}}\left(x_n^{(+)}\right) + \bar{N}_{\mathrm{hl}}\left(x_n^{(-)}\right)\right\}\Big], \tag{2.15}$$

if N_{R} is odd, and

$$\frac{\bar{G}_{\mathrm{H}}^{(\mathrm{p})}(x_0)}{k_{\mathrm{B}}T} = \frac{1}{N_{\mathrm{R}}} \sum_{n=1}^{\frac{1}{2}N_{\mathrm{R}}} \left\{ \frac{\bar{G}_{\mathrm{H}}\left(x_n^{(+)}\right)}{k_{\mathrm{B}}T} + \frac{\bar{G}_{\mathrm{H}}\left(x_n^{(-)}\right)}{k_{\mathrm{B}}T} \right\}, \tag{2.16}$$

$$\bar{N}_{\rm hl}^{\rm (p)}(x_0) = \frac{1}{N_{\rm R}} \sum_{n=1}^{\frac{1}{2}N_{\rm R}} \left\{ \bar{N}_{\rm hl}\left(x_n^{(+)}\right) + \bar{N}_{\rm hl}\left(x_n^{(-)}\right) \right\}, \tag{2.17}$$

if $N_{\rm R}$ is even. The function $\bar{N}_{\rm hl}(x)$ and the backbone contribution to the function $\bar{G}_{\rm H}(x)$ that enter these expressions are those used in Sec. 2.3 for profiles of long polypeptides. We now present results for two cases of this scenario. Results for two cases of this scenario follow in Sec. 2.4.3 below.

2.4.2 Scenario #2

Here we consider free-energy landscapes for short peptides of varying position and orientation. The model peptide in this scenario consists of two straight segments as illustrated in Fig. 2.11. In applications to pHLIP, the position of the kink could naturally be associated with the helix inhibiting proline residue but it can be anywhere in the sequence. One simulation study [10] places a kink at the position of the Asp residue which is somewhat closer to the N terminus than the Pro residue.

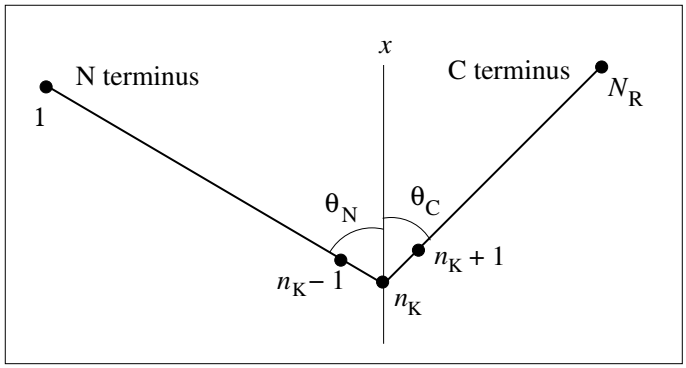

Figure 2.11. Variables $x_{n_{\rm K}}$ (depth of kink) and $\theta_{\rm N}$, $\theta_{\rm C}$ (angles of adjacent segments) used in scenario #2.

We use the position $x_{n_{\rm K}}$ of the kink on the normal to the bilayer and the angles $\theta_{\rm N}$, $\theta_{\rm C}$ of the segments with the N, C termini, respectively, as the variables that uniquely specify the position and orientation of the peptide in the membrane

environment. The distances between residues in each segment are modeled as before but with all positions of residues now generated recursively from the position of the kink.

Free-energy landscapes that depend on three continuous variables are hard to visualize in a compact fashion. We must adapt the mode of graphical presentation to best serve our chief purpose, which is to explore different pathways of insertion such as have been suggested, for example, on the basis of experimental evidence. We therefore investigate the variation of free energy along different insertion pathways. Each pathway is described by a specific synchronized variation of the variables indicated in Fig. 2.11. In practice, we express all three variables as functions of a single parameter with range $0 \leq z \leq 1$. Each set of functions, $x_{n_\mathrm{K}}(z)$, $\theta_\mathrm{N}(z)$, $\theta_\mathrm{C}(z)$, traces a specific insertion pathway.

In pHLIP applications the initial state (named state II) is an adsorbed peptide, largely oriented parallel to the bilayer and positioned close to the lipid headgroups. The final state (named state III) is the TM state, oriented perpendicular to the bilayer, with the C terminus having crossed the membrane. Hence the initial and final values of the three variables are

$$x_{n_\mathrm{K}}(0) = x_{n_\mathrm{K}}^{\mathrm{ini}}, \quad \theta_\mathrm{N}(0) = \theta_\mathrm{C}(0) = \frac{\pi}{2}, \tag{2.18a}$$

$$x_{n_\mathrm{K}}(1) = x_{n_\mathrm{K}}^{\mathrm{fin}}, \quad \theta_\mathrm{N}(1)0, \quad \theta_\mathrm{C}(1) = \pi, \tag{2.18b}$$

where the values $x_{n_\mathrm{K}}^{\mathrm{ini}}$, $x_{n_\mathrm{K}}^{\mathrm{fin}}$ are the results of optimizations. Along each insertion pathway we can calculate the free energy as we have done in Sec. 2.4.1. Modified versions of (2.12), (2.14), (2.15) read

$$x_{n-1} = x_n + l(x_n)\cos\theta_\mathrm{N} \quad : \; n = n_\mathrm{K}, \ldots, 2, \tag{2.19}$$

$$x_{n+1} = x_n + l(x_n)\cos\theta_\mathrm{C} \quad : \; n = n_\mathrm{K}, \ldots, N_\mathrm{R} - 1,$$

$$\frac{G_\mathrm{H}^{(\mathrm{K})}(x_{n_\mathrm{K}})}{k_\mathrm{B}T} = \sum_{n=1}^{N_\mathrm{R}} \frac{\bar{G}_\mathrm{H}(x_n)}{k_\mathrm{B}T}, \tag{2.20}$$

$$N_{\rm hl}^{(\rm K)}(x_{n_{\rm K}}) = \sum_{n=1}^{N_{\rm R}} \bar{N}_{\rm hl}(x_n). \tag{2.21}$$

For the function $l(x_n)$, which enters (2.19), we again refer to Sec.2.4.3 below, where two variants are being used. Evidently, scenario #1 is contained in scenario #2 as a special case.

A stretch of pathway along which the free energy as calculated via (2.20) decreases is favorable for insertion and a stretch along which it increases is unfavorable. Insertion pathways that are all downhill are fast. Any barriers encountered slow down the characteristic time of insertion.

The free-energy densities $\bar{G}_{\rm H}(x)$ from Sec. 2.3 include only selected contributions, mostly from the backbone of the peptide. We therefore postpone applications of scenario #2 until we are ready to employ amended expressions $\bar{G}_{\rm H}(x)$ that include the most important side chain contributions.

2.4.3 Landscapes from backbone alone

It is instructive to investigate the backbone contributions to the landscapes of helicity and free energy for scenario #1. The simplest case uses

$$l(x) = l_{\rm h} = 1.5\text{Å}, \tag{2.22}$$

implying that the position coordinate between successive residues progresses uniformly and independently of conformation. The value chosen in (2.22) is most realistic for a helical segment.

Free-energy and helicity landscapes thus predicted are shown in Fig. 2.12 for $\mu = 2, \infty$ and three size of peptides ($N_{\rm R} = 35, 47, 23$). The first row of landscapes ($N_{\rm R} = 35$) is tailored to represent some variant of pHLIP, the second row a significantly longer longer peptide ($N_{\rm R} = 47$), and the third row a significantly shorter peptide ($N_{\rm R} = 23$). We only consider enthalpy parameters that are neutral $\alpha_{\rm H} = 1$ or represent a gain $\alpha_{\rm H} = 1.05, 1.1$.

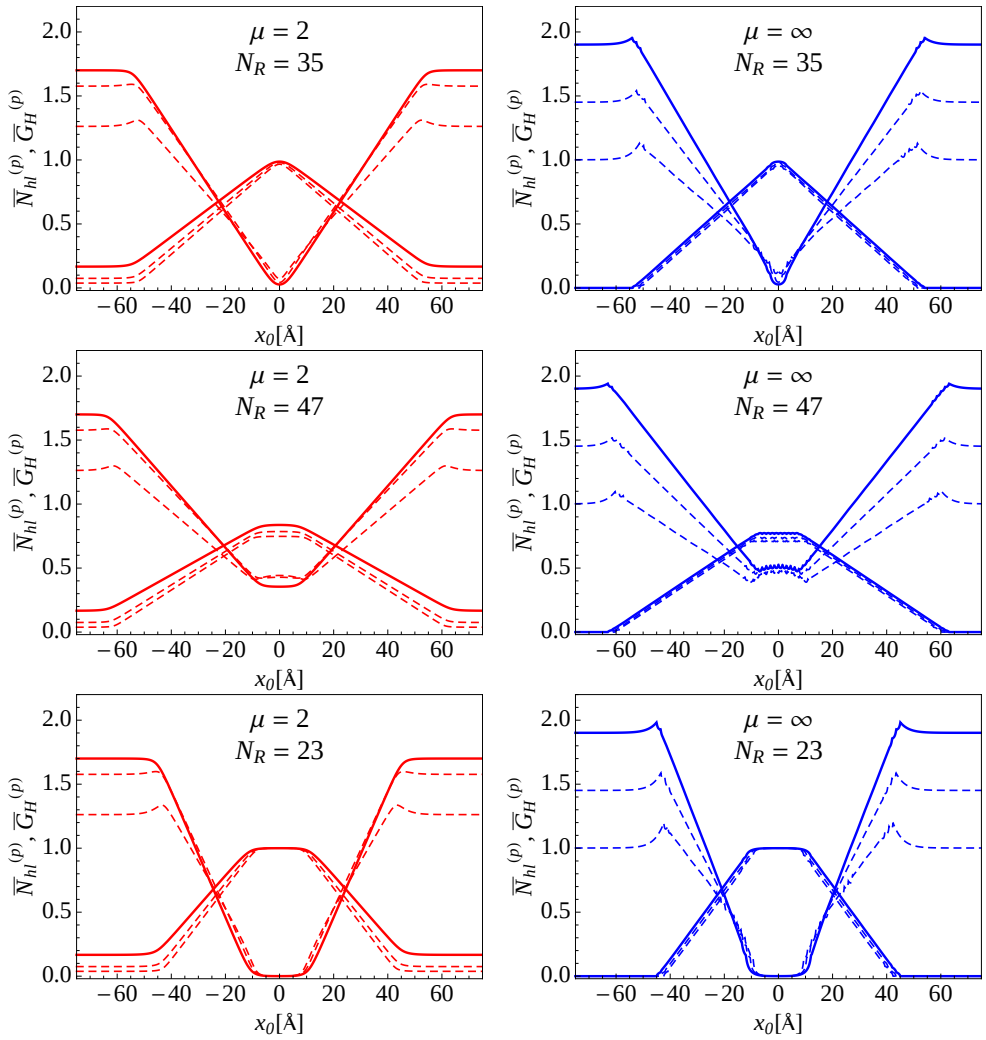

Figure 2.12. Helicity (peaked at center) and scaled free-energy of peptide with $N_{\rm R}$ residues oriented and positioned as described in the text versus the coordinate x_0 of the central residue for $\mu = 2$ (left) and $\mu = \infty$ (right) and three sizes. The solid and dashed curves pertain to the values $\alpha_{\rm H} = 1$ and $\alpha_{\rm H} = 1.05, 1.1$ of the enthalpy parameter, respectively.

No matter whether we assign little entropy to coil segments ($\mu = 2$) or a lot of entropy ($\mu = \infty$), the landscapes have similar shapes. The helicity landscapes are almost independent of the enthalpy parameter. That parameter affects the free-energy primarily in the aqueous environment as expected.

Insertion into the membrane is clearly favored in all three cases and for both variants of the model. The plots also tell us that insertion is accompanied by a conformational change from coil to helix. For the longest peptide the minimum in free energy is not as deep and the maximum in helicity is not as high as is the case for the two shorter ones. The obvious reason is that the former has significant flanking ends that remain in water.

Of particular interest is the free energy barrier that separates states with the center of the peptide in aqueous or membrane environments, the former mostly in coil conformation and the latter in helix conformation. This free-energy barrier is very shallow for $\mu = 2$ and only exists if $\alpha_{\mathrm{H}} > 1$. For $\mu = \infty$, on the other hand, it is more conspicuous and present even for $\alpha = 1$. This difference is related to the higher entropy that must be expelled by coil segments for $\mu = \infty$ when they order into helix segments before they can cash in the enthalpic benefit of the lipid environment.

One message we take from this simplest scenario is that insertion is not automatic. An environmental change may be needed to push the peptide over the barrier. One environmental change that acts on protonatable charges carried by some side chains (e.g. Asp and Glu) is an increase in acidity. The consequences for pHLIP are well documented by experiments [11, 12, 13, 14]. Let us recall that the extent of insertion as predicted by free-energy landscapes and the extent of ordering as predicted by helicity landscapes can be directly monitored experimentally, namely by tryptophan fluorescence and by circular dichroism experiments,

respectively.

An improved level of modeling takes into account that the distance between adjacent residues is different in the coil and helix conformations. This feature can readily be implemented in the framework of our methodology. Here we keep the orientation of the peptide perpendicular to the bilayer. The distance between successive residues now depends on the local conformation of the backbone at their position in the membrane environment. That conformation is either coil or helix with probabilities for which we use propensity profile as calculated in Sec. 2.3. We replace (2.22) by

$$\begin{aligned} \Delta x \doteq & x_{n+1} - x_n = l(x_n), \\ & l(x_n) = l_{\mathrm{c}} - (l_{\mathrm{c}} - l_{\mathrm{h}})\bar{N}_{\mathrm{hl}}(x_n), \end{aligned} \tag{2.23}$$

where we use (averaged) lengths l_{c} for coil segments and l_{h} for helix segments. The values of l_{c} and l_{h} are two physical parameters in this model version.

We again mark the location of the peptide as the position of its center residue or center link and generate the relative positions of residues recursively from that center as in (2.12) and (2.13). The expression for free energy per residue and for helicity do not change from (2.14)-(2.16) in structure.

Whereas the shift l_{h} associated with a helix segment remains fixed at the value (2.22) our model now allows for a coil segment of average length l_{c} that can be different. Its maximum value is the contour length per residue of the backbone: $l_{\mathrm{c}} \lesssim 4\text{Å}$.

In Figs. 2.13 and 2.14 we show landscapes of the helicity and the free energy for the cases $l_{\mathrm{c}} = 2.0\text{Å}$ and $l_{\mathrm{c}} = 3.0\text{Å}$, respectively. Many of the features that we described in the context of Fig. 2.12 remain qualitatively the same. However, the improved model adds one new feature to the landscapes of both helicity and free energy. It is a change in slope at the interface position. With l_{c} increasing, the

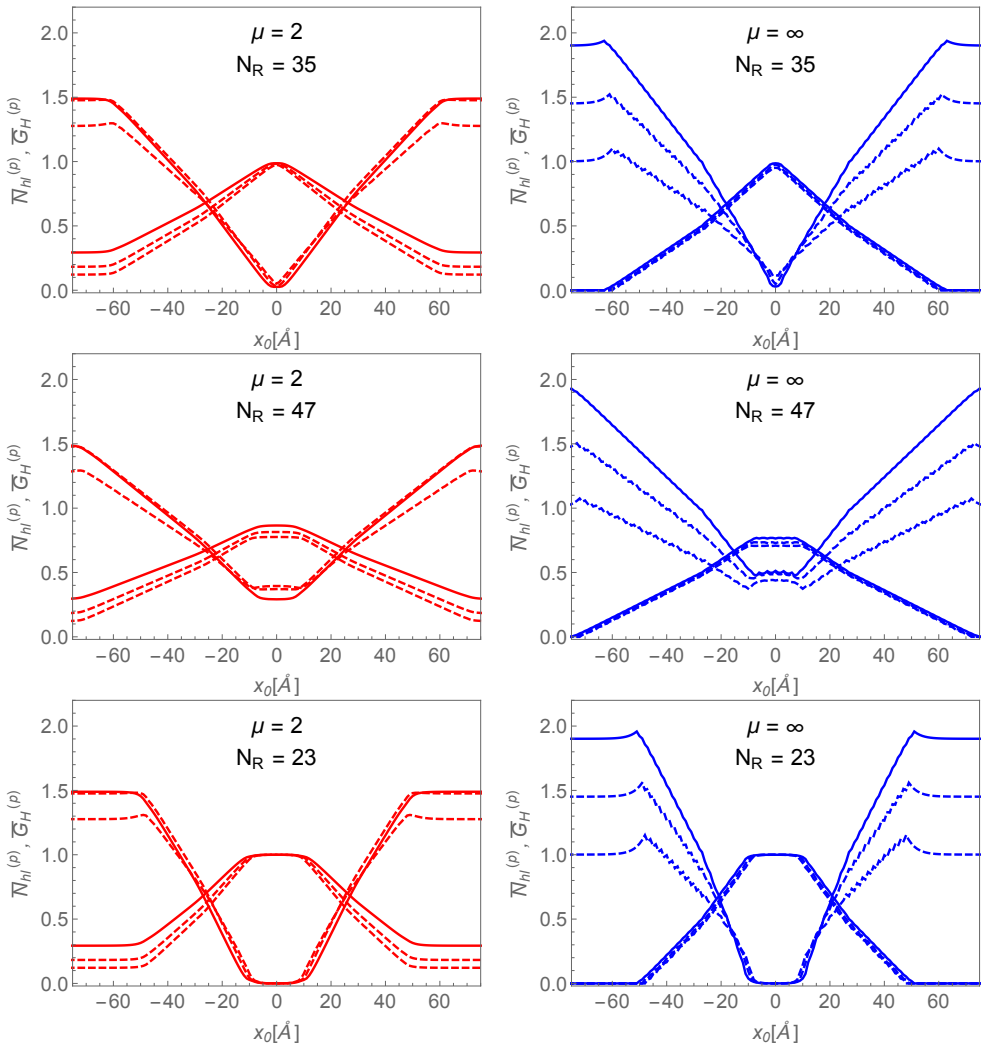

Figure 2.13. Helicity (peaked at center) and scaled free-energy of peptide with $N_{\rm R}$ residues oriented and positioned as described in the text versus the coordinate x_0 of the central residue for $\mu = 2$ (left) and $\mu = \infty$ (right) and three sizes. The solid and dashed curves pertain to the values $\alpha_{\rm H} = 1$ and $\alpha_{\rm H} = 1.05, 1.1$ of the enthalpy parameter, respectively. The length specifications are $l_{\rm h} = 1.5$Åand $l_{\rm c} = 2.0$Å.

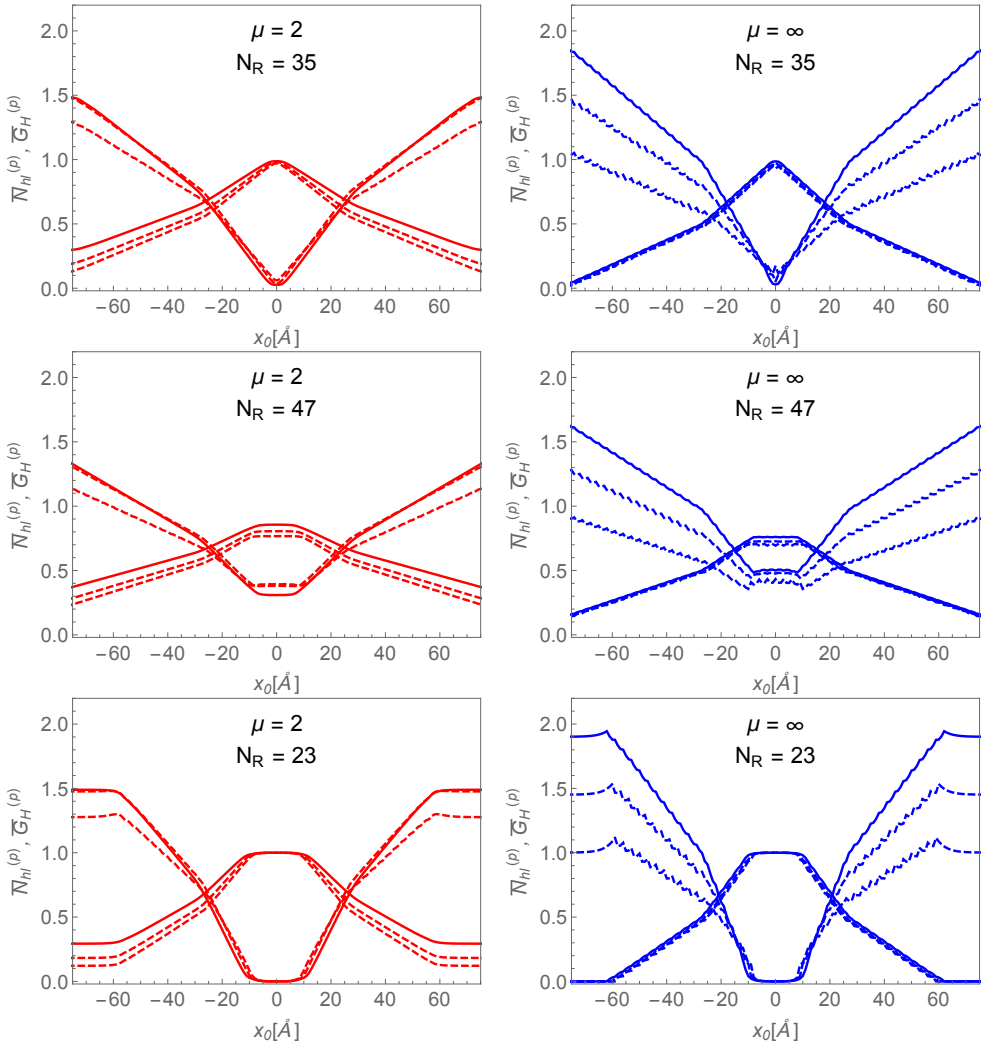

Figure 2.14. Helicity (peaked at center) and scaled free-energy of peptide with $N_{\rm R}$ residues oriented and positioned as described in the text versus the coordinate x_0 of the central residue for $\mu = 2$ (left) and $\mu = \infty$ (right) and three sizes. The solid and dashed curves pertain to the values $\alpha_{\rm H} = 1$ and $\alpha_{\rm H} = 1.05, 1.1$ of the enthalpy parameter, respectively. The length specifications are $l_{\rm h} = 1.5$Åand $l_{\rm c} = 3.0$Å.

bend becomes more pronounced. In addition we observe a systematic change in a feature already present in Fig. 2.12. The locations where both landscapes level off shift further away from the center of them membrane.

Consider the the peptide with $N_{\mathrm{R}} = 35$ residues. It just fits into the non-polar space between the two layers of headgroups. As its center residue is gradually displaced away from the middle of the membrane the two stretches of residues on either side respond differently two the changing environment. The outer stretch gradually moves into the polar environment, converting from helix to coil in the process. The inner half remains in helix conformation. The outer stretch gets continually longer a while the inner stretch maintains its length.

For as long as the center residue is inside the membrane thus observe a steady descent of the helicity and a steady rise of the free energy. When the center residues moves into the non-polar environment, then the inner stretch begins to unravel from helix to coil and thus increase is length. The implication is the helical portion of the inner stretch now moves away from the center of the membrane at a slower rate than the center residue does. The consequences are that the both landscapes now change more slowly and level off later.

The systematic dependence on N_{R} of the landscapes remain very similar in the improved model as is evident when we compare the panels of Fig. 2.12 with the corresponding panels of Figs. 2.13 and 2.14.

2.5 Conclusion and outlook

The continuation of this project will focus on the effects of side chains and additional effects of the lipids. The contributions to the free-energy landscape originating from the side chains of a peptide with a given sequence of residues are manifold. Some of them are associated with properties of the lipid bilayer that we have yet to discuss. For this investigation the protagonist will be a peptide

with a specific sequence of residues, in particular the pHLIP variants identified in Fig. 2.2.

We shall estimate the peptide free energy (in the sense explained in Sec. 2.2.2) in three principal configurations relative to its environment and along specific pathways between two of them. The three states are

- State I: pHLIP is in aqueous solution and in coil conformation.
- State II: pHLIP is adsorbed to the outside interface of a cell membrane with interstitial fluid or of a liposome with water. At high pH the degree of adsorption is shallow and the conformation is coil. At low pH the degree of adsorption is deeper and the conformation is α-helix, at least to a large part.
- State III: pHLIP is in a trans-membrane state with a helical central part and short coil-like flanking ends.

Given that the side chains of residues range from strongly hydrophobic to strongly hydrophilic, it is not surprising that they contribute significantly to the free-energy landscapes. Some side chains carry (positive or negative) electric charges. The protonation of the negatively charged Glu and Asp residues and of the negatively charged C terminus at low pH changes the overall hydrophobicity significantly. The current consensus is that the protonations caused by a drop in pH trigger the destabilization of the adsorbed state II in favor of the trans-membrane state III.

For the continuation of our modeling we plan to employ and compare existing data for transfer free energies in order to estimate the side-chain contribution to the free-energy owing to hydrophobicity. Our goal is to compare such free-energy contributions between states I, II, and III at high pH and at low pH.

The primary effect of the lipids has been present in this study from the outset.

The lipids produce the heterogeneous membrane environment described in Sec. 2.2 and govern the conformational propensities of peptides placed into this environment as described in Sec. 2.3. In the projected continuation we are concerned with secondary effects of the lipids as they interact with pHLIP in its adsorbed or trans-membrane states. Our focus will be on three such effects that appear most relevant for this study:

- lateral pressure profile within lipid bilayer,
- hydrophobic mismatch of peptide in trans-membrane state,
- lipid entropy reduction due to contact with peptide

The first two effects are enthalpic in nature and the third entropic. It is not a priori clear which effects are the most important. Further effects may require consideration.

List of References

[1] G. P. Sharma, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, "Coil-helix transition of polypeptide at water-lipid interface," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, p. P01034, 2015.

[2] G. A. Nemnes and D.-V. Anghel, "Fractional exclusion statistics in non-homogeneous interacting particle systems," *Roman. Rep. Phys.*, vol. 66, no. 336, 2014.

[3] C. Moore, D. Liu, B. Ballnus, M. Karbach, and G. Müller, "Disks in narrow channel jammed by gravity and centrifuge: profiles of pressure," *J. Stat. Mech*, vol. 2014, p. P04008, 2014.

[4] B. Bakhti, M. Maass, M. Karbach, and G. Müller, "Monodisperse rods in external potentials," *Phys. Rev. E*, vol. 92, p. 042112, 2015.

[5] L. Wang, "Measurements and implications of the membrane dipole potential," *Annual Review of Biochemistry*, vol. 81, pp. 615–635, 2012.

[6] C. Hofsäß, E. Lindahl, and O. Edholm, "Molecular dynamics simulations of phospholipid bilayers with cholesterol," *Biophysical Journal*, vol. 84, no. 4, pp. 2192–2206, 2003.

[7] J. L. MacCallum, W. D. Bennett, and D. P. Tieleman, "Distribution of amino acids in a lipid bilayer from computer simulations," *Biophysical Journal*, vol. 94, no. 9, pp. 3393–3404, 2008.

[8] A. V. Finkelstein and O. Ptitsyn, *Protein Physics*. AIP Press, New York, 2002.

[9] A. Grosberg and A. Khokhlov, *Statistical Physics of Macromolecules*. AIP Press, New York, 1994.

[10] D. Vila-Viçosa, T. F. Silva, G. Slaybaugh, Y. K. Reshetnyak, O. A. Andreev, and M. Machuqueiro, "Membrane-induced pka shifts in wt-phlip and its l16h variant," *Journal of Chemical Theory and Computation*, vol. 14, no. 6, pp. 3289–3297, 2018.

[11] Y. K. Reshetnyak, M. Segala, O. A. Andreev, and D. M. Engelman, "A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers," *Biophysical Journal*, vol. 93, no. 7, pp. 2363–2372, 2007.

[12] Y. K. Reshetnyak, O. A. Andreev, M. Segala, V. S. Markin, and D. M. Engelman, "Energetics of peptide (phlip) binding to and folding across a lipid bilayer membrane," *Proceedings of the National Academy of Sciences*, vol. 105, no. 40, pp. 15 340–15 345, 2008.

[13] D. Weerakkody, A. Moshnikova, M. S. Thakur, V. Moshnikova, J. Daniels, D. M. Engelman, O. A. Andreev, and Y. K. Reshetnyak, "Family of ph (low) insertion peptides for tumor targeting," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5834–5839, 2013.

[14] O. A. Andreev, A. D. Dupuy, M. Segala, S. Sandugu, D. A. Serra, C. O. Chichester, D. M. Engelman, and Y. K. Reshetnyak, "Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo," *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7893–7898, 2007.

CHAPTER 3

Kinetic Model for Peptide Insertion into a Membrane I: Kinetics of Protonatable Residue at Varying pH

3.1 Introduction

The folding mechanism of water-soluble proteins including the coil-helix transition has been thoroughly investigated over many years. The coil-helix transition that accompanies the insertion of a polypeptide into a membrane has not nearly been studied as extensively theoretically and experimentally. The theoretical modeling is much more complex due to (i) the heterogeneous environment in which the conformational changes occur and (ii) the simultaneity of insertion and conformational change. Experimental studies are limited to the small selection of polypeptides that are both water soluble and exhibit controlled spontaneous insertion/folding and exit/unfolding.

Stability and folding of all membrane proteins, irrespective of the mechanism of their insertion into the membrane, are governed by the formation of a secondary structure in the lipid bilayer environment, driven by the hydrophobic effect and by hydrogen bonding. Consideration of these factors has led to the proposition that the spontaneous insertion and folding includes the formation of helical intermediates at the bilayer surface, followed by insertion [1, 2, 3, 4, 5]. The evidence from fluorescent experiments is mostly indirect [6, 7, 8] and the results of simulation studies are somewhat ambiguous in this respect [5, 9, 10, 11, 12, 13, 14, 15].

The kinetic modeling presented in this work is custom-made, as were two predecessor studies [16, 17] that serve as its foundation, for variants of the *pH Low Insertion Peptide* (pHLIP) family but not to the exclusion of other peptides with similar attributes. All variants of pHLIP considered here are moderately hydrophobic membrane peptides containing protonatable residues. Our published

biophysical data indicate that at neutral and high pH, pHLIP is monomeric and largely unstructured [18].

Peptides in aqueous solution (state I) coexist with peptides adsorbed to the surface of a lipid bilayer (state II). The fraction of the adsorbed peptides is controlled by the lipid-to-peptide ratio. Lowering the pH shifts the equilibrium towards membrane insertion and the formation of a trans-membrane (TM) helix (state III). A subsequent increase of pH initiates a process of unfolding of the TM helix and exit from the bilayer (states I or II). The three way stations of pHLIP realized at thermal equilibrium are schematically illustrated in Fig. 3.1.

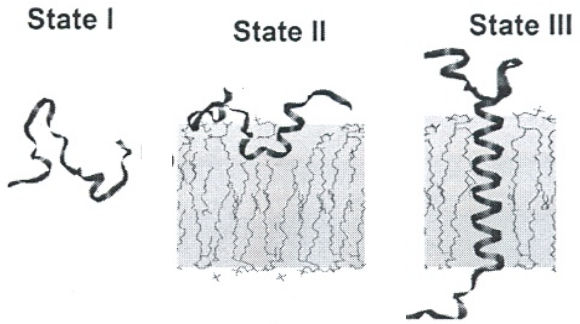

Figure 3.1. Schematic presentation of the three pHLIP positions relative to the lipid bilayer and the associated conformation at thermal equilibrium for specific peptide concentrations and levels of pH.

The process of the peptide association with the membrane is experimentally distinguishable from the process of peptide partitioning into bilayer. The latter is accompanied by the coil-helix transition and triggered by a drop in pH. Caloric experiments [19, 20] have established that the bilayer affinity of the peptide is about 30-50 times higher at low pH than at high pH.

Another fact is that pHLIP insertion comes in the wake of the protonation of Asp/Glu residues in the trans-membrane (TM) part and the inserting end (C terminus) of the peptide. The protonation leads to an increase in hydrophobicity which, in turn, triggers the peptide to partition into the bilayer and fold across

the lipid bilayer [21, 22]. Since the peptide is located at the border between polar (aqueous) and non-polar (membrane) environments, the pK_a for the protonation of Asp and Glu residues is significantly shifted to higher values [23]. Moreover, the pK_a of individual groups changes during peptide propagation into a bilayer [18, 20, 22, 24, 25, 26].

In Sec. 3.2 we describe the working hypothesis underlying our kinetic modeling, first in general terms and then regarding specific tasks ahead. The physical phenomena to be described involve a combination of processes on a range of time scales and with interactions that affect these time scales. For molecular dynamics (MD) simulations this is a serious challenge. Our Markov chain modeling is designed to cope with this challenge. It can do that at a price that is significant (spatial resolution) but still worth the effort.

Next we present Markov chain models for two separate processes. One process describes changes in protonation status of one protonatable residue at a given location in the membrane environment with a controllable level of pH in water (Sec. 3.3). The other process describes changes in location of that residue (in a given protonation status) across the membrane environment as driven by the prevailing forces (Sec. 3.4).

The two component processes are combined, as reported in Sec. 3.5, by way of an interlocking Markov chain process. Each component process has an intrinsic time scale that can be varied parametrically. The time scale of the combined process depends on these parameter choices directly by changing the speed of one or the other process and indirectly via natural feedback mechanisms.

In Sec. 3.6 we sketch out extensions of the modeling already carried out to include multiple sites with protonatable charges that are linked by backbone segments in either coil or helix conformation. Changes in conformation would then

be a third Markov chain process. The dynamics of conformational change depends on location. Motion, in turn, depends on protonation status. A Markov chain process that interlocks all three component processes is the logical next step of kinetic modeling.

3.2 Working Hypothesis

The proposed modeling of the peptide insertion is largely anchored by experimental data on pHLIP.

3.2.1 Experimental evidence

The initial state in the kinetic model is a pHLIP molecule adsorbed at the surface of a bilayer in coil conformation. The first step in the insertion process, taking place on a $\sim$100ms time scale, is associated with a coil-helix transformation [27].

The second step (insertion, formation of TM helical state) would be completed within the same time scale if no protonatable residues or polar cargo were present at the peptide inserting end (C terminus) [28]. The presence of protonatable residues (or polar cargo) leads to the stabilization of intermediate states. These intermediate states involve helical segments located near the water-lipid interface and oriented mostly parallel to it.

There are two dominant effective forces: a hydrophobic force $\mathbf{F}_{\text{in}}$ pointing toward the center of the bilayer and an electric force $\mathbf{F}_{\text{out}}$ pointing away from the bilayer. The force $\mathbf{F}_{\text{out}}$ acts directly on the negative charges of the Asp/Glu residues while the force $\mathbf{F}_{\text{in}}$ originates from nearby non-polar residues (e.g. Leu, Ile). It is well established that the removal of Asp/Glu residues from the peptide inserting end increases the rate of insertion by orders of magnitude, thus greatly reducing the stopping power of $\mathbf{F}_{\text{out}}$ [28].

3.2.2 Forces in control of insertion

Consider a coil segment near one of the Asp residues. At pH 8 the Asp residue has a high probability for being in a deprotonated (negatively charged) state. There are two major effective forces acting on that segment: a hydrophobic force $\mathbf{F}_{\text{in}}$ pointing toward the lipids (positive x) and an electric force $\mathbf{F}_{\text{out}}$ pointing away from the lipids (negative x) (Fig. 3.2). The force $\mathbf{F}_{\text{out}}$ mostly originates from negatively charged Asp residues. The force $\mathbf{F}_{\text{in}}$ originates from nearby hydrophobic residues (Leu and Ile, for example). Both forces depend on the heterogeneity of the medium, its shifting attribute from polar to non-polar with increasing position coordinate x. We expect both forces to have profiles with maxima (in strength) near the interface, and tails fading in both directions.

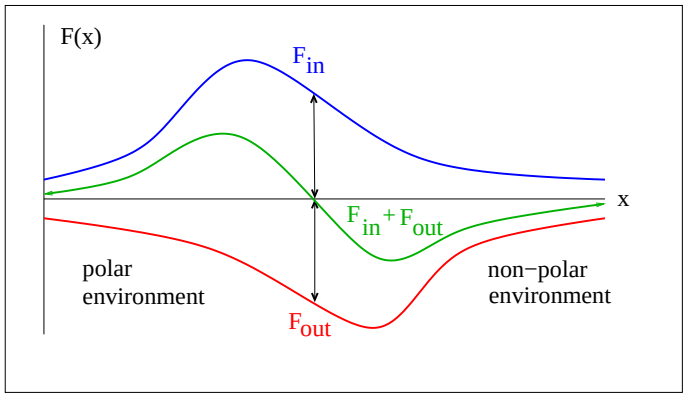

Figure 3.2. Schematic profile of the hydrophobic force $\mathbf{F}_{\text{in}}$, the electric force $\mathbf{F}_{\text{out}}$, and the sum. The position $x = 0$ is at the outer water-lipid interface.

For a stable equilibrium near the interface to exist, for which there is ample experimental evidence at high pH, the maximum of $\mathbf{F}_{\text{in}}$ must be further out than the maximum of $\mathbf{F}_{\text{out}}$ as sketched. This establishes a potential minimum at the position of the vertical arrows. Near this mechanical equilibrium position the force $\mathbf{F}_{\text{in}} + \mathbf{F}_{\text{out}}$ acts like a restoring force.

The average state of protonation of the Asp residue is determined by its pK_a in

a given environment. In aqueous solution the pK_a of protonation of Asp residues is around 4. As the environment changes from polar to non-polar across the interface, the pK_a of protonation rises. Hence the probability of protonation at constant pH increases while the Asp residue moves into the lipid environment.

The force imbalance caused by protonation of the Asp residue when the pH is dropped initiates its displacement toward a location where deprotonation becomes less likely. Deprotonation, if it happens in time, restores the balance of forces. The threshold of an instability that leads to the insertion then hinges on a comparison of two time scales. One is related to the rates of protonation/deprotonation at a given location in the heterogeneous environment and at a given level of pH. The other is the time needed to displace the Asp a critical distance under the influence of the unbalanced hydrophobic force $\mathbf{F}_{\rm in}$ while it stays protonated.

3.2.3 Kinetics of protonatable residue

There are strong experimental indicators that the insertion of a pHLIP molecule from (adsorbed) state II to the (transmembrane) state III is driven by the kinetics of protonatable residues near the C terminus. In the modeling presented in the following we focus on one such residue and consider two processes that overlap in time. We first analyze them separately, in Secs. 3.3 and 3.4, and then, in Sec. 3.5, we combine them. They are

- a process of protonation/deprotonation of Asp/Glu residues controlled by the availability of protons and the polar/nonpolar nature of the environment,
- a process of in/out motion through the heterogeneous membrane environment driven by unbalanced hydrophobic and electrostatic forces.

The speed and direction of each process depends on the instantaneous status of the other process.

We begin with the design of Markov chain models for each process, which we then combine into a single stochastic process. The challenge we face is to simultaneously accommodate all relevant time scales. In a Markov chain model one universal clock ticks at the rate in which transition matrices are multiplied. The physical time scales are controlled via adjustments in transition rates.

3.3 Protonation status of residue

We begin with a negatively charged, protonatable residue such as Glu or Asp in water. If the rate constants of deprotonation and protonation are $k_{\rm d}$ and $k_{\rm p}$, respectively, then chemical equilibrium in water dictates (as detailed in Appendix C) that the ratio depends on the level of pH and the $\rm pK_a$ of the residue as follows:

$$\frac{k_{\rm d}}{k_{\rm p}} = 10^{\rm pH-pK_a}. \tag{3.1}$$

If the residue is located in the heterogeneous membrane environment, i.e. away from pure water, then the pH in (3.1) must be replaced by $\frac{1}{2}\rm pK_w$, where $\rm pK_w$ is the pK of water in that environment. In Appendix C we present further details about the modification of (3.1). We conclude that between pure water and the center of the membrane both $k_{\rm d}$ and $k_{\rm p}$ decrease with $k_{\rm d}$ decreasing faster than $k_{\rm p}$. In the following modeling we introduce parameters that reflect key attributes of the rate constants $k_{\rm p}$, $k_{\rm d}$.

3.3.1 Markov chain

The Markov chain model discretises the line of positions between pure water and the center of the membrane into an array of N cells (see Fig. 3.3). For convenience we assign odd i and even $i+1$ to the same position of the protonatable residue in its protonated (pro) and deprotonated (dep) states, respectively. The index $i = 1, \ldots, 2N$ thus specifies the position of the residue and its pro/dep status

[1].

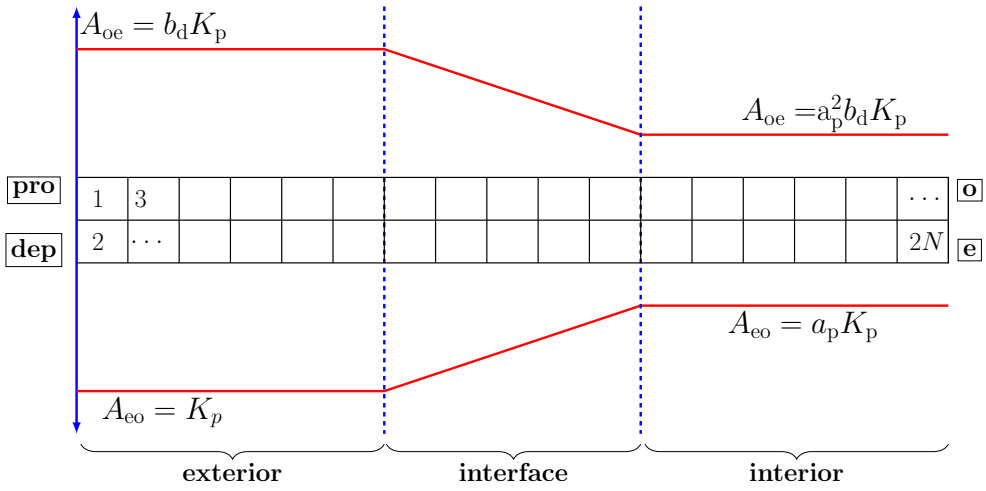

Figure 3.3. Off-diagonal matrix elements $A_{\rm eo}$ and $A_{\rm oe}$, where e stands for the even index and o for the odd index at the same position. The parameter $a_{\rm p}$ is controlled by the change in $\mathrm{pK_a}$ of the protonatable residue. The parameter $b_{\rm d}$ is controlled by the level of pH. The parameter $K_{\rm p}$ sets the time unit of the computational clock.

The Markov chain model for the pro/dep kinetics at a fixed position is characterized by a transition matrix of the form,

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & 0 & 0 & 0 & \cdots \\ A_{21} & A_{22} & 0 & 0 & 0 & \cdots \\ 0 & 0 & A_{33} & A_{34} & 0 & \cdots \\ 0 & 0 & A_{43} & A_{44} & 0 & \cdots \\ 0 & 0 & 0 & 0 & A_{55} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.2}$$

The physics is contained in the nonzero off-diagonal elements $A_{\rm eo}$ and $A_{\rm oe}$ between states with odd index i and even index $i+1$. The diagonal elements are determined by the requirement that probability is normalized:

$$\left.\begin{aligned} A_{i,i} &= 1 - A_{i,i+1} \\ A_{i+1,i+1} &= 1 - A_{i+1,i} \end{aligned}\right\} \;:\; i = 1, 3, \ldots 2N-1. \tag{3.3}$$

[1] A mathematically more elegant notation would employ a rank-2 tensor to specify the state and a rank-4 transition tensor. However, our emphasis on elucidating competing time scales is better with the alternative notation.

3.3.2 Transition rates

For the specification of the transition probabilities of protonation and deprotonation as encoded in the matrix elements $A_{\rm eo}$ and $A_{\rm oe}$, respectively, we divide the range of positions into three intervals of variable relative size, representing the exterior, interface, and interior parts of the membrane, respectively (see Fig. 3.3). Across the interface, both transition probabilities change between the values,

$$A_{\rm eo} = \begin{cases} K_{\rm p} & : \text{ ext.} \\ a_{\rm p} K_{\rm p} & : \text{ int.} \end{cases} \qquad A_{\rm oe} = \begin{cases} b_{\rm d} K_{\rm p} & : \text{ ext.} \\ a_{\rm p}^2 b_{\rm d} K_{\rm p} & : \text{ int.} \end{cases} \tag{3.4}$$

The drop of one rate by a factor $a_{\rm p}$ and the other by a factor of $a_{\rm p}^2$, where $0 < a_{\rm p} < 1$, encodes the slowing-down of both processes between polar and non-polar environments. The linear variation of $A_{\rm eo}$ and $A_{\rm oe}$ across the interface as indicated in Fig. 3.3 is the default choice. It can be modified on the basis of empirical evidence.

The rationale behind our choices of parametrization for these rates is explained in Appendix C. The significance of the three model parameters are the following:

- $K_{\rm p}$ sets the time scale for the pro/dep process associated with the computational clock. The latter advances by one tick per matrix multiplication in the Markov chain model.
- $b_{\rm d}$ is controlled by the level of pH.
- $a_{\rm p}$ is controlled by the change in $\rm pK_a$ (and $\rm pK_w$) between water and membrane interior.

3.3.3 Time scale

With all ingredients for the kinetics of the protonation status in place we must choose an initial state as encoded in a (normalized) probability vector,

$$\mathbf{P}(0) = \Big(P_1(0), P_2(0), \ldots, P_{2N}(0) \Big). \tag{3.5}$$

It specifies the initial probability for the residue being at a given position in a given protonation status. Here we are concerned with the time evolution of the pro/dep status alone. As mentioned previously, on the computational clock time advances one tick per matrix multiplication, expressed as follows:

$$\mathbf{P}(n) = \mathbf{P}(n-1) \cdot \mathbf{A} = \mathbf{P}(0) \cdot \mathbf{A}^n. \tag{3.6}$$

For the Markov chain process specified by matrix $\mathbf{A}$ all times t will be stated in units of K_p^{-1}. This has two advantages.

- By choosing the value of K_p sufficiently small we can ensure that the computational clock provides the necessary time resolution of the underlying physical process.
- For values of K_p below a certain threshold, the results expressed as functions of t become independent of K_p.

3.3.4 Stationary state and equilibration

The structure of the matrix $\mathbf{A}$ guarantees that a stationary state $\mathbf{P}(\infty)$ exists. The block-diagonal nature of $\mathbf{A}$ conserves probability within each pair of states,

$$\left.\begin{aligned} P_i(n) &= P_\mathrm{o}(n), \\ P_{i+1}(n) &= P_\mathrm{e}(n), \end{aligned}\right\} \;:\; i = 1, 3, \ldots 2N-1. \tag{3.7}$$

The stationary state can then be computed iteratively,

$$\mathbf{P}_\mathrm{s} = \mathbf{P}(\infty) = \lim_{n\to\infty} \mathbf{P}(0) \cdot \mathbf{A}^n, \tag{3.8}$$

or determined from the eigenvalue equation,

$$\mathbf{P}_\mathrm{s} = \mathbf{P}_\mathrm{s} \cdot \mathbf{A}, \tag{3.9}$$

and the normalization as an auxiliary condition. We find

$$P_\mathrm{s}^{(\mathrm{o})} = \frac{A_\mathrm{eo}}{A_\mathrm{eo} + A_\mathrm{oe}}, \quad P_\mathrm{s}^{(\mathrm{e})} = \frac{A_\mathrm{oe}}{A_\mathrm{eo} + A_\mathrm{oe}}. \tag{3.10}$$

For the purpose of illustrating the local equilibration of the protonation status for situations where the water is at high pH or low pH we consider a model system with $N = 7$ positions. The main point is to get a grip on the variable time scale associated with the pro/dep process in different environments. The border between exterior and interface is at the position represented by indices 5,6 and the border between interface and interior by indices 9,10. All elements of the matrix $\mathbf{A}$ are determined by the parameters $K_{\rm p}$, $b_{\rm d}$, and $a_{\rm p}$ as explained in Sec. 3.3.2.

Our choices of parameter values for a situation of high pH (in water) and a situation of low pH produce transition rates and stationary states as illustrated in Figs. 3.4(a) and 3.4(b), respectively.

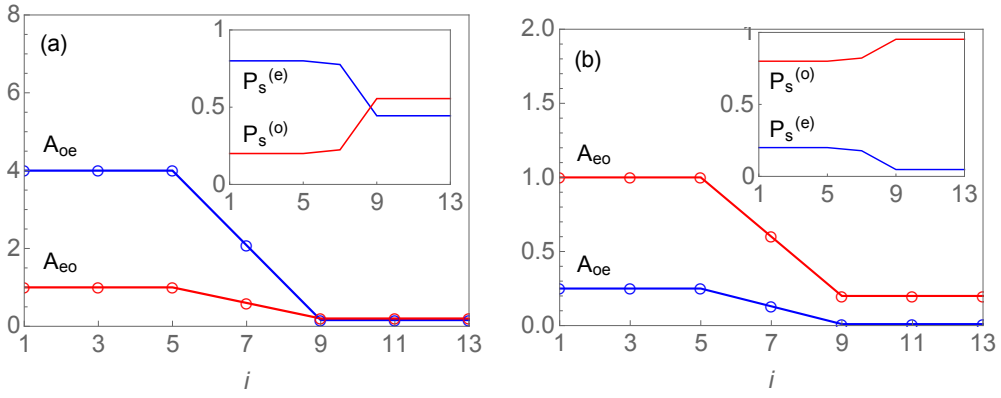

Figure 3.4. Scaled off-diagonal elements $A_{\rm eo}/K_{\rm p}$ and $A_{\rm oe}/K_{\rm p}$ at the $N = 7$ positions, here represented by indices $i = 1, 3, \ldots, 13$. The parameter values in panel (a), $b_{\rm d} = 4$, $a_{\rm p} = 0.2$, are designed to represent a situation of high pH, and the parameter values in panel (b), $b_{\rm d} = 0.25$, $a_{\rm p} = 0.2$, a situation of low pH. The insets represent the stationary state (3.10) for each case.

3.3.5 High level of pH

For this case, the deprotonation rate $A_{\rm oe}$ in water is much higher than the protonation rate $A_{\rm eo}$, yielding, as we shall see, a high probability for the residue to be deprotonated. Both rates slow down as the position shifts from the exterior to the interior region. The deprotonation rate slows down more rapidly as

argued in Appendix C. In Fig. 3.5 we show the dependence on physical time t of the probabilities P_i, P_{i+1} with opposite initial values for positions in the exterior, interface, and interior regions. The value of K_p was chosen sufficiently small to ensure optimal time resolution.

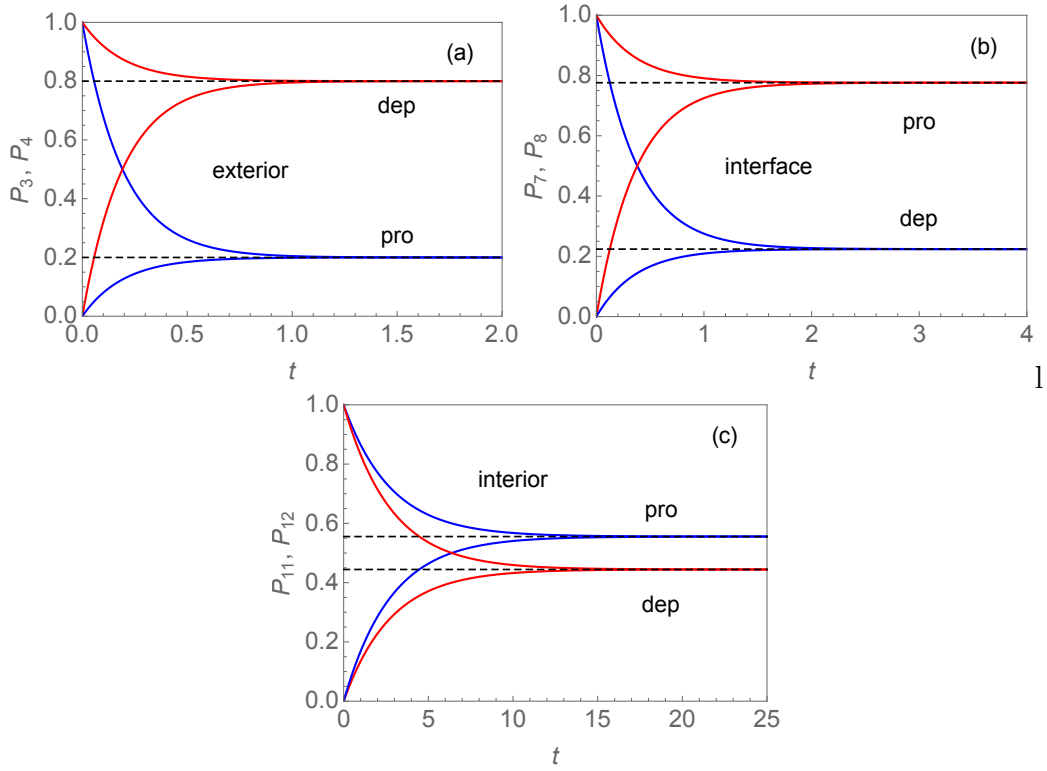

Figure 3.5. Equilibration of probabilities P_3, P_4 in the exterior region, P_7, P_8 at the center of the interface, and P_{11}, P_{12} in the interior region at a high level of pH (in water). The parameters are specified in Fig. 3.4(a). The initial state for each probability shown is either fully protonated or fully deprotonated. All curves were produced with $K_\mathrm{p} = 0.01$. The dashed lines represent the equilibrium values from (3.10).

Taking into account the different horizontal scales we can clearly see the progressive slowing down of the pro/dep process as we move from the exterior region, Fig. 3.5(a), across the interface, Fig. 3.5(b), to the interior, Fig. 3.5(c). The time it takes to reach equilibrium is much the for opposite initial states. We can also observe how the rising $\mathrm{pK_a}$ between exterior and interior regions changes the

equilibrium values (3.10). At high pH, the equilibrium in water strongly favors deprotonation. That tendency becomes much weaker as the position shifts from the polar exterior to the non-polar interior.

3.3.6 Low level of pH

Here we are concerned with the time evolution of the pro/dep process at constant, low pH. For that purpose we have lowered the relevant parameter from $b_{\mathrm{d}} = 4$ to $b_{\mathrm{d}} = 0.25$. The parameter a_{p} is left unchanged. The modified transition rates at low pH are shown in Fig. 3.4(b) along with the stationary state (3.10).

We see that the lower pH affects the rate of deprotonation, A_{oe}, more strongly than the rate of protonation, A_{eo}. The former is now much lower both in water and inside the membrane. This shifts the equilibrium toward the protonated state everywhere.

The equilibration process at the same positions as in Fig. 3.4 with the same initial states but now in an environment of lower pH produces the results shown in Fig. 3.6. In water (exterior region) the equilibrium has shifted from favoring the deprotonated state at high pH to favoring the protonated state at low pH. The time scale for equilibration has not changed significantly.

Inside the membrane (interior region), the protonated state was already favored over the deprotonated state at high pH albeit not by much. At low pH it is much more highly favored. The time scale of equilibration remains slower in the membrane compared to the water environment. In summary, the change in pH does not alter the speed of equilibration significantly, but the change in environment does.

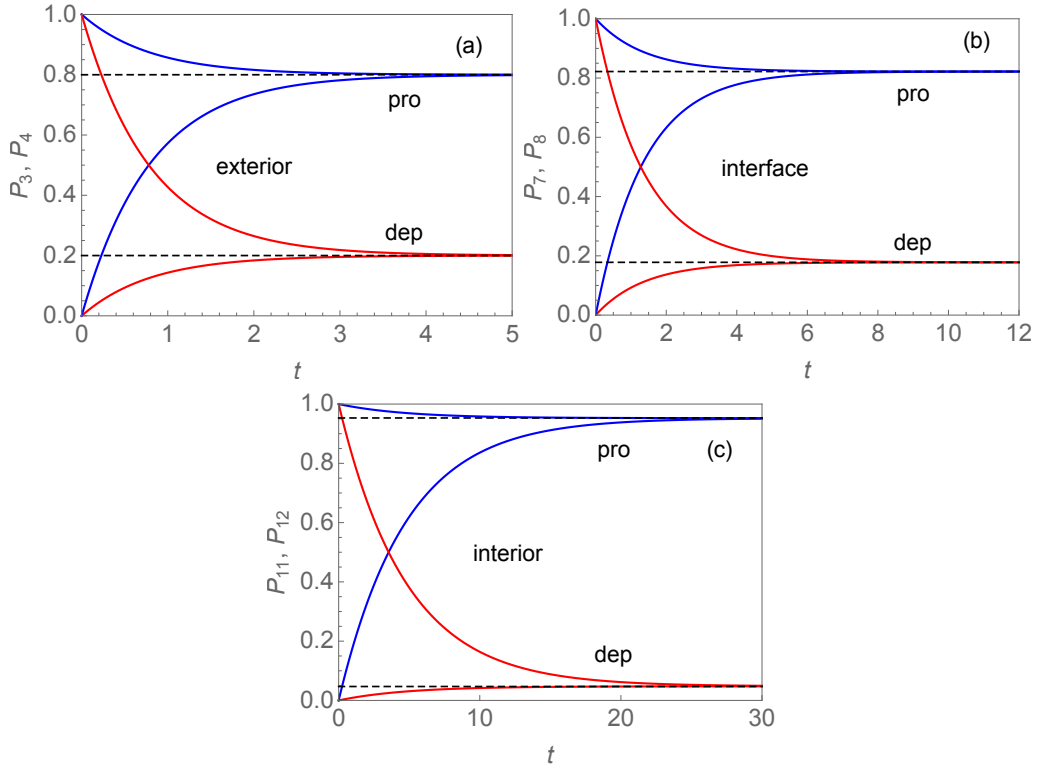

Figure 3.6. Equilibration of probabilities P_3, P_4 in the exterior region, P_7, P_8 at the center of the interface, and P_{11}, P_{12} in the interior region at a high level of pH (in water). The parameters are specified in Fig. 3.4(a). The initial state for each probability shown is either fully protonated or fully deprotonated. All curves were produced with $K_{\mathrm{p}} = 0.01$. The dashed lines represent the equilibrium values from (3.10).

3.3.7 Drop of pH

Lowering the pH as part of a pHLIP insertion experiment involves the mixing of fluids, which takes place on its own time scale. The so-called dead time before the first data are taken is of the order of 10ms, which is certainly much longer than the time scale on which the protonation status equilibrates, at least, in water. Equilibration is slower in the membrane environment but, in all likelihood, comes to completion well within the dead time. This conclusion will be used in Sec. 3.5, where we combine the kinetics of protonation status and motion.

3.4 Motion of residue

Depending on whether the residue is protonated or deprotonated it experiences different forces. We have already identified these forces in Sec. 3.2.2 as being predominantly hydrophobic and elecrostatic in nature with profiles as sketched in Fig. 3.2.

It is reasonable to argue that the translocation of the protonatable residue is an overdamped motion driven by these forces. The velocity is then proportional to the magnitude of the resultant force. In the Markov chain model, the units of both distance and time are fixed. The velocity of motion is accounted for by the probability of a step of unit distance in ($i \to i+2$) or out ($i \to i-2$) in the time period assigned to one matrix multiplication (computational clock).

3.4.1 Transition rates

The motion of the residue with fixed protonation status is encoded in the transition matrix,

$$\mathbf{B} = \begin{pmatrix} B_{11} & 0 & B_{13} & 0 & 0 & \cdots \\ 0 & B_{22} & 0 & B_{24} & 0 & \cdots \\ B_{31} & 0 & B_{33} & 0 & B_{35} & \cdots \\ 0 & B_{42} & 0 & B_{44} & 0 & \cdots \\ 0 & 0 & B_{53} & 0 & B_{55} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{3.11}$$

where in each row we have, in addition to the nonzero diagonal element $B_{i,i}$, one nonzero off-diagonal element, $B_{i,i-2}$ if the resultant force is directed out, or a nonzero $B_{i,i+2}$ if the resultant force is directed in. The value of the nonzero $B_{i,i\pm2}$ is proportional to the magnitude of the resultant force. The values of the diagonal elements again follow from the normalization condition,

$$B_{i,i} = 1 - B_{i,i-2} - B_{i,i+2}. \tag{3.12}$$

The direction and magnitude of the resultant force depends on the position and pro/dep state of the residue. The force is $\mathbf{F}_{\text{in}}$ if the residue is protonated and

$\mathbf{F}_{\rm in} + \mathbf{F}_{\rm out}$ if it is deprotonated.

As a default, we use piecewise linear force profiles with $\mathbf{F}_{\rm in}$ and $\mathbf{F}_{\rm out}$ related to each other by two successive reflections as illustrated in Fig. 3.7. This choice, which can be modified as prompted by empirical evidence, produces a stable equilibrium at the center of the membrane when only $F_{\rm in}$ is acting as is the case when the residue is protonated. When both forces are acting as is the case when the residue is deprotonated, then they produce a stable equilibrium state at the center of the interface.

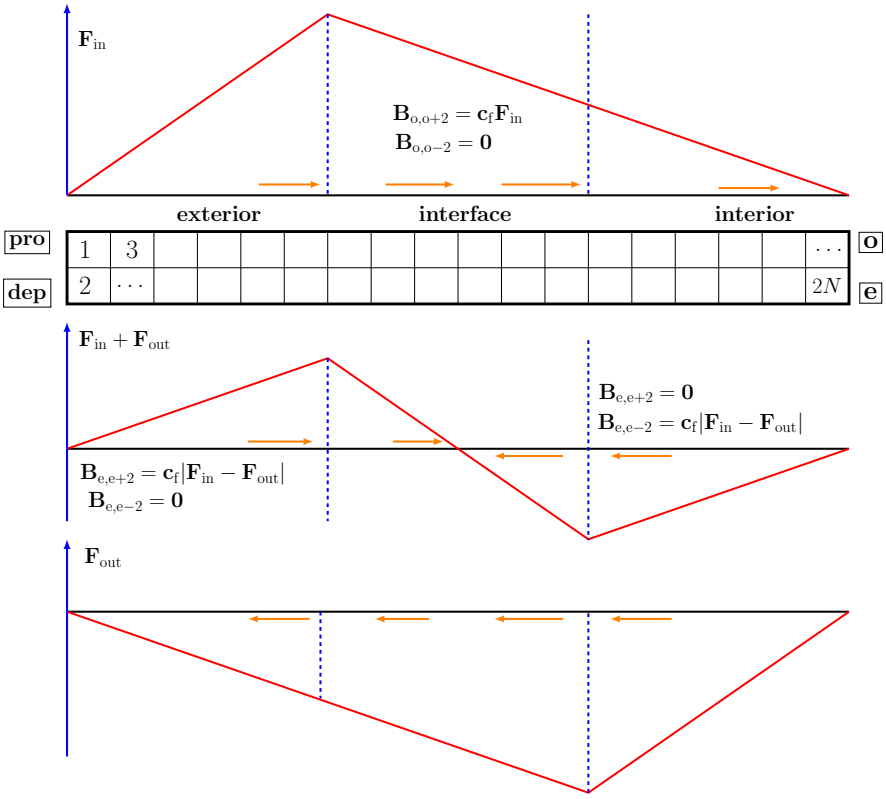

Figure 3.7. Piecewise linear force profiles for hydrophobic force $\mathbf{F}_{\rm in}$ and electrostatic force $\mathbf{F}_{\rm out}$. They determine the off-diagonal elements $B_{\rm e,e\pm2}$ and $B_{\rm o,o\pm2}$ as indicated. The parameter $c_{\rm f}$ sets the time unit of the computational clock.

This simple model force profile leaves us with a single model parameter $c_{\rm f}$.

We write

$$B_{\mathrm{o,o+2}} = c_{\mathrm{f}}\bar{F}_{\mathrm{in}}, \quad B_{\mathrm{o,o-2}} = 0, \tag{3.13a}$$

$$B_{\mathrm{e,e+2}} = \begin{cases} c_{\mathrm{f}}|\bar{F}_{\mathrm{in}} - \bar{F}_{\mathrm{out}}| & : \ \bar{F}_{\mathrm{in}} > \bar{F}_{\mathrm{out}}, \\ 0 & : \bar{F}_{\mathrm{in}} < \bar{F}_{\mathrm{out}}, \end{cases} \tag{3.13b}$$

$$B_{\mathrm{e,e-2}} = \begin{cases} 0 & : \bar{F}_{\mathrm{in}} > \bar{F}_{\mathrm{out}}, \\ c_{\mathrm{f}}|\bar{F}_{\mathrm{in}} - \bar{F}_{\mathrm{out}}| & : \ \bar{F}_{\mathrm{in}} < \bar{F}_{\mathrm{out}}, \end{cases} \tag{3.13c}$$

where $\bar{F}_{\mathrm{in}} = F_{\mathrm{in}}/F_{\max}$ and $\bar{F}_{\mathrm{out}} = F_{\mathrm{out}}/F_{\max}$ are scaled forces.

3.4.2 Time scale and stationarity

Equations (3.8) and (3.9) with matrix **B** substituted for **A** determine the stationary state. The time unit associated with one tick of the computational clock, i.e. one matrix-**B** multiplication as in (3.6), is controlled by the parameter c_{f}. When we express the physical time t in units of c_{f}^{-1} the results again become independent of the parameter c_{f} and provide optimal time resolution provided we choose a value below a certain threshold.

The stationary states are readily identified by inspection of the force profiles in Fig. 3.7. Protonated residues will end up in the interior (state $i = 2N - 1$) and deprotonated residues on the interface (state $i = N$), irrespective of the initial conditions. These are absorbing states of matrix **B**, in the Markov chain jargon [2].

In the following we separately investigate the motion of protonated and deprotonated residues as driven by the applicable forces. The protonated residue only experiences the hydrophobic force $\mathbf{F}_{\mathrm{in}}$. The deprotonated residue also experiences the electrostatic force $\mathbf{F}_{\mathrm{out}}$.

[2] If the residue is protonated then the unique absorbing state is $i = 2N - 1$. There are transitions in but no transitions out. If the residue is protonated then the state $i = N$ is the unique absorbing state for even $N/2$ whereas the states $i = N, N + 2$ are two absorbing states for odd $N/2$.

3.4.3 Protonated residue

Locations of the residue in its protonated state have odd indices i. The curves in Fig. 3.8(a) represent the time evolution of the probabilities P_i, $i = 1, 3, \ldots, 2N-1$, if at time $t = 0$ the residue is located in the exterior region (state $i = 1$).

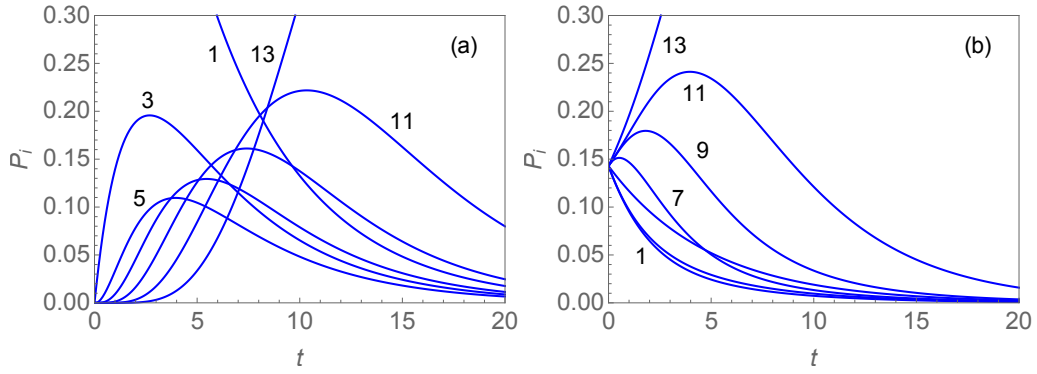

Figure 3.8. Probability versus time of protonated residue during motion driven by force $\mathbf{F}_{\text{in}}$ for two sets of initial conditions: (a) the residue is initially at position $i = 1$, (b) the residue is initially at any position with equal probability. All curves were produced with $c_{\text{f}} = 0.1$.

P_1 is monotonically decreasing as t increases and will approach zero asymptotically for $t \to \infty$. The P_i for intermediate positions $i = 3, 5, \ldots, 11$ are represented (in that order) by the curves that start at zero and approach a smooth maximum at later and later times. All these curves also reach zero asymptotically for $t \to \infty$. P_{13} also starts from zero but is monotonically increasing and will level off at 1 asymptotically for $t \to \infty$. The non-monotonic variation in the heights of the maxima reflects information about average position and velocity as will be analyzed below.

In Fig. 3.8(b) we have only changed the initial state. We now assume that the (protonated) residue is in any position with equal probability. All P_i except P_{13} reach zero asymptotically for $t \to \infty$. The residue again ends up with unit probability at the center of the membrane. The curves for low values of i are

monotonically decreasing whereas the curves for higher i go through a maximum. This feature encodes information about the unidirectional motion to be analyzed later.

3.4.4 Deprotonated residue

Locations of the residue in its deprotonated state have even indices i. The motion is now driven by two opposing forces. The time evolution of the P_i, $i = 2, 4, \ldots, 2N$ shown in Fig. 3.9(a) are for the case where at $t = 0$ the residue is at the outermost position (state $i = 2$) with unit probablity. By contrast, the data in Fig. 3.9(b) represent the case where the residue is at each position with equal probability. We can see in the two panels that the residue moves toward the center of the interface, here located at position $i = 8$.

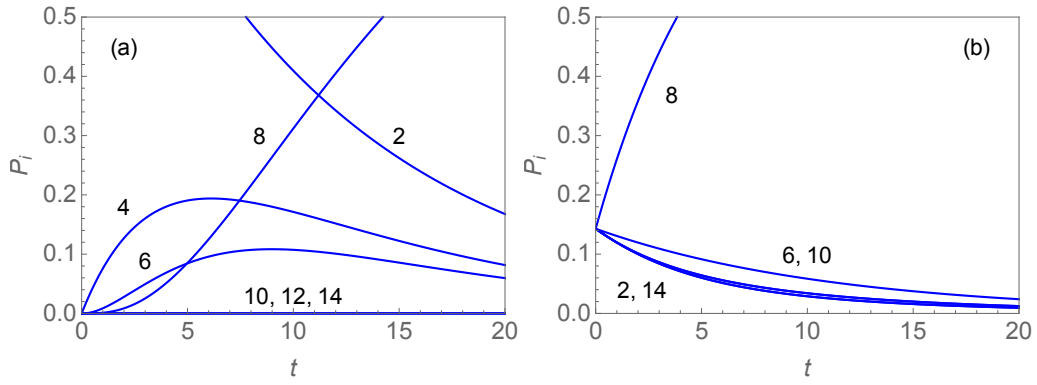

Figure 3.9. Probability versus time of deprotonated residue during motion driven by force $\mathbf{F}_{\rm in} + \mathbf{F}_{\rm out}$ for two sets of initial conditions: (a) the residue is initially at position $i = 2$, (b) the residue is initially at any position with equal probability. All curves were produced with $c_{\rm f} = 0.1$.

In panel (a) P_{10}, P_{12}, and P_{14} stay strictly zero. The probability P_2 is monotonically decreasing and the probability P_8 monotonically increasing between the same extreme values. The probabilities P_4, and P_6 start from zero and go back to zero asymptotically. The initial rise is quicker and higher for the former. The long-time asymptotics is the same for both. In panel (b) the reflection symmetry

of the combined forces makes the curves overlap in pairs. Naturally, the final state is the same in both panels.

3.4.5 Location and speed

Some of the information contained in Figs. 3.8 and 3.9 can be unlocked by using different formats. For example, if our chief interest is in the location of the residue and the speed at which it changes location under given circumstances we can proceed as follows.

We consider the cases where the initial state is either $i = 1$ or $i = 2$ as represented by Figs. 3.8(a) and 3.9(a). Both residues can be at N distinct scaled positions across the range $0 \leq x_i \leq 1$. The position x_i is related to the (odd or even) state index of the protonated or deprotonated residue via

$$x_i = \begin{cases} \dfrac{i-1}{N-2} & : \text{odd } i, \\ \dfrac{i-2}{N-2} & : \text{even } i. \end{cases} \tag{3.14}$$

The average position $\langle x \rangle$ of the protonated and the deprotonated residue as functions of physical time t is then calculated as different weighted averages of state probabilities.

The results are shown as two curves in Fig. 3.10. The protonated residue, which experiences only the hydrophobic force $\mathbf{F}_{\text{in}}$, moves to the center of the membrane ($x = 1$) whereas the deprotonated residue, which is also subject to the electrostatic force $\mathbf{F}_{\text{out}}$, moves to the interface ($x = 0.5$).

Both residues are released in water ($x = 0$) at time $t = 0$. They reach their destination within 30 units of physical time. That time unit is arbitrary in the sense that it can be assigned any value based on empirical evidence. We have seen that changing the value of c_{f} below a certain threshold leaves the shape of the curves on the time scale t invariant.

In the overdamped dynamics used here the residues inertia is neglected, whcih

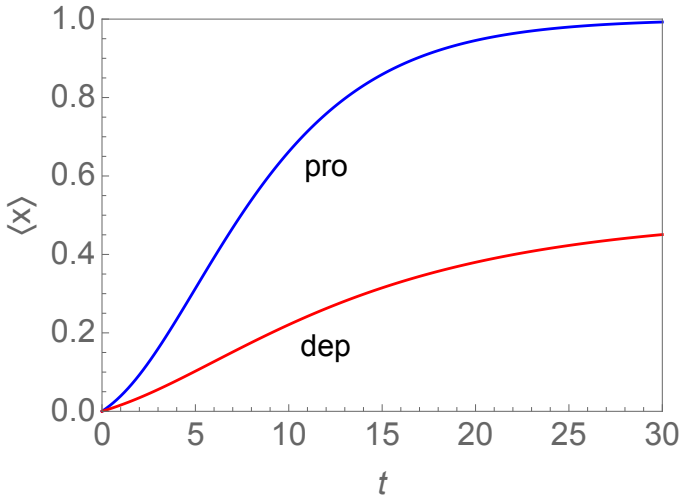

Figure 3.10. Mean scaled position versus time of a protonated and a deprotonated residue when released in water $\langle x \rangle = 0$. The interface is centered at at $x = 0.5$. The center of the membrane is at $\langle x \rangle = 1$. The physical time is $t = n \cdot c_\mathrm{f}$, where n is the number of ticks on the computational clock and we choose $c_\mathrm{f} = 0.1$ for optimized time resolution and computational cost.

explains the nonzero average initial velocity as indicated by the slope of each curve. The protonated residue experiences a stronger force than the deprotonated residue. In consequence the former moves faster and reaches its destination a bit earlier even though it travels a longer distance.

3.5 Motion with status change

In Secs. 3.3 and 3.4 we have explored two processes that we intend to combine here. The physical time scale relative to the tick of the computational clock was controllable by a single parameter in each case. That parameter was K_p for matrix **A** and c_f for matrix **B**. In both processes we achieved results whose t-dependence were independent of those parameter values provided they were chosen sufficiently small. The unit of t in each process was arbitrary and could be assigned any value based on empirical evidence.

One way to combine the two processes within the framework of the same

methodology employs the transition matrix

$$\mathbf{W} = \mathbf{A} \cdot \mathbf{B} \tag{3.15}$$

with $\mathbf{A}$ from (3.2) and $\mathbf{B}$ from (3.11). Whereas the t-dependence of Markov chain process driven by matrix $\mathbf{A}$ alone or by matrix $\mathbf{B}$ alone are independent of the relevant parameters K_p and c_f, respectively, provided they are sufficiently small, the t-dependence of the Markov chain process driven by matrix $\mathbf{W}$ varies with the ratio $r_\mathrm{w} \doteq c_\mathrm{f}/K_\mathrm{p}$. A small (large) value of that ratio means that the kinetics of motion is slow (fast) compared to the kinetics of protonation and deprotonation. Variations of r_w are likely to produce very diverse results as will be explored in the following.

The dynamics of $\mathbf{W}$ has a sort of adiabatic regime. For small r_w all motion takes place with the pro/dep status always close to being equilibrated. In practice this means that the motion consists of tiny steps forward and back that average out to an effectively smooth motion of a residue with fractional protonation status. If we release the residue at some location then its subsequent motion does not significantly depend on its initial protonation status. This adiabatic limit is expected to produce the least complex dynamics.

As we investigate the dynamics of the process $\mathbf{W}$ for progressively larger values of r_w, the motion becomes dependent on the initial protonation status. The residue is now able to move a significant distance, at least in some locations, before the protonation status has a chance to switch.

From Sec. 3.3 we know that the protonation status shifts its equilibrium as a function of position and the equilibrium value in water is controllable by the level of pH. From Sec. 3.4 we know that the deprotonated residue tends to settle at the interface and the protonated residue in the center of the membrane.

With the level of pH we now control the average initial protonation status

with which the residue is being released in water. It is to be expected that the residue will either settle at the interface (adsorbed state) or in the center of the membrane (inserted state) with probabilities that depend on the level of pH.

3.5.1 Stationarity

It makes sense to begin our exploration with stationary states, i.e. solutions of the eigenvalue equation,

$$\mathbf{P}(\infty) = \mathbf{P}(\infty) \cdot \mathbf{W}. \tag{3.16}$$

These solutions are all independent of the initial state of the residue, both regarding location and protonation status. We are looking for the physically relevant regions in the space spanned by the three parameters a_{p} (controlling the change in $\mathrm{pK_a}$ between exterior and interior), b_{d} (controlling the pH in the exterior), and r_{w} (controlling the relative time scales of processes **A** and **B**).

The solution of (3.16) for given values of these three parameters produces a unique stationary probability distribution,

$$\mathbf{P}(\infty) = \left(P_1^{(\mathrm{s})}, \ldots, P_{2N}^{(\mathrm{s})}\right), \tag{3.17}$$

containing all information about that state. In Fig. 3.11 we show stationary distributions for for a model with $N = 7$, where we only vary parameter b_{d}.

Panel (a) represents a situation of high pH. The protonatable residue is with very high probability in state $i = 8$. This state represents a deprotonated residue at the center of the interface. As we decrease b_{d} in the three steps we observe the probabilities to shift towards locations in the interior and then also toward protonation. In panel(d) the residue is with overwhelming probability in state $i = 13$, meaning protonated and located at the center of the membrane.

As we explore the parameter space in search of the physically relevant and interesting regions, it is advisable to focus on quantities that are graphically more

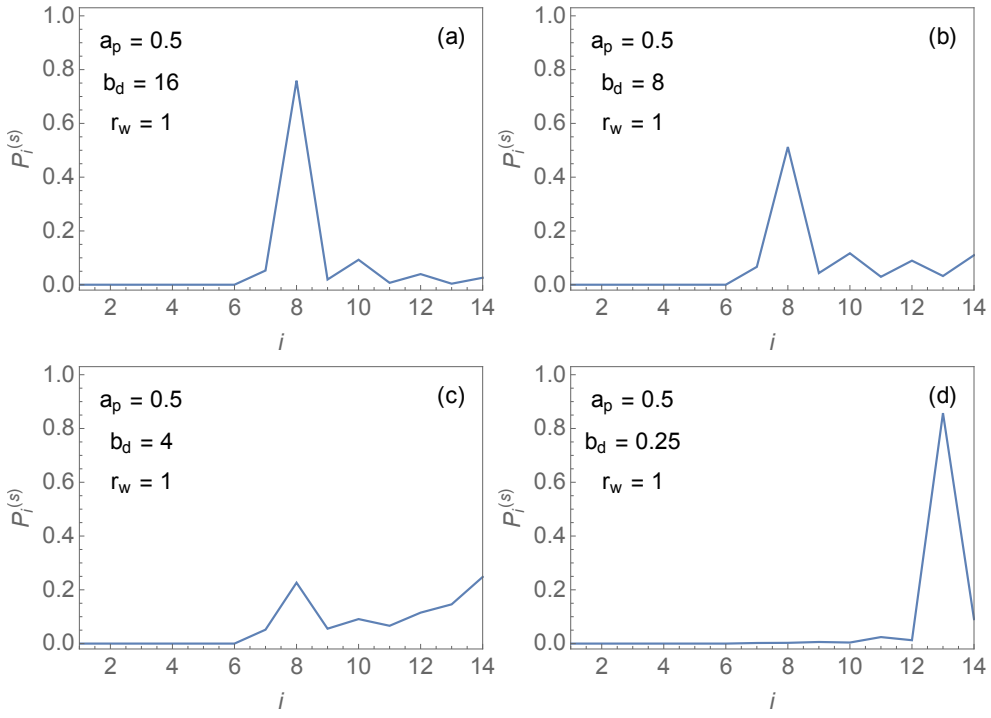

Figure 3.11. Stationary probability distribution $P_i^{(\mathrm{s})}$, $i = 1, \ldots, N$ for $N = 14$. Note the variations in the parameter b_d controlling the pH in water.

compact than the $P_i^{(\mathrm{s})}$, yet express the most important information. Two natural selections are the mean scaled position of the residue,

$$\langle x \rangle^{(\mathrm{s})} \doteq \sum_{i=1}^{N} x_i P_i^{(\mathrm{s})}, \tag{3.18}$$

with x_i from (3.14), and the probabilities

$$P_{\mathrm{pro}}^{(\mathrm{s})} = \sum_{\text{odd } i} P_i^{(\mathrm{s})}, \quad P_{\mathrm{dep}}^{(\mathrm{s})} = \sum_{\text{even } i} P_i^{(\mathrm{s})}, \tag{3.19}$$

for its protonation status. The latter are, of course, complementary: $P_{\mathrm{pro}}^{(\mathrm{s})} + P_{\mathrm{dep}}^{(\mathrm{s})} = 1$. In Fig. 3.12 we have plotted the three quantities versus the parameter b_d across the same range but now varied continuously. We see that a quasistatic rise in pH lowers the protonation probability and shifts the average position of the residue from near the center of the membrane, $\langle x \rangle^{(\mathrm{s})} = 1$, to near the center of the interface, $\langle x \rangle^{(\mathrm{s})} = 0.5$.

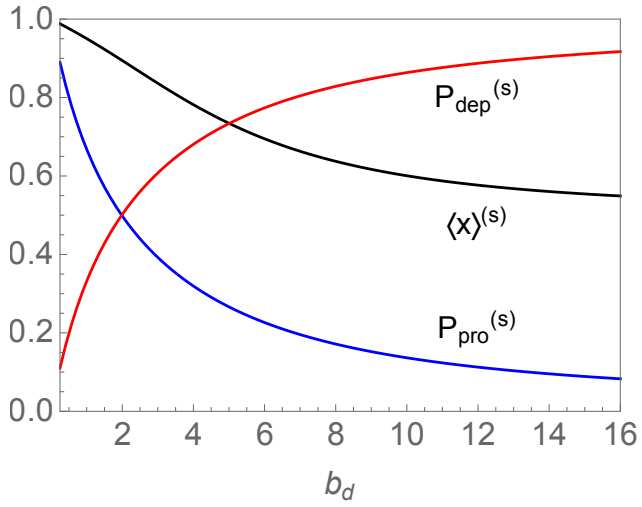

Figure 3.12. Mean scaled position $\langle x \rangle^{(s)}$ and probabilities $P_{pro}^{(s)}$, $P_{dep}^{(s)}$ for the protonation status of the residue versus parameter $0.25 \leq b_d \leq 16$. All other parameters are as in Fig. 3.11.

How are these stationary states affected when we vary the other two parameters a_p and r_w? This is shown in Fig. 3.13. In panel (a) we vary a_p. Recall that the lowest value of a_p represents the highest rise in pK_a of the residue between exterior and interior positions. We observe that variations in that parameter have little impact at low pH (small b_d). The effect is largest at intermediate pH and still significant at high pH. Residues with a large shift in pK_a tend to be located deeper in the membrane at high pH than residues with a smaller shift. The former tend to have a higher protonation probability than the latter. The second trend is, of course, at least in part, a consequence of the first trend.

By strong contrast, but unsurprisingly, a variation in r_w produces much weaker changes as illustrated in panel (b) for just one case. Decreasing r_w from unity by as much as a factor of 10 (not shown) has little effect. Increasing r_w by the same factor produces the effect shown. Recall that changing r_w means changing the relative time scales between the two intertwined processes **A** and **B**. A factor of 10 either way is huge, yet we have to keep in mind that here we are studying

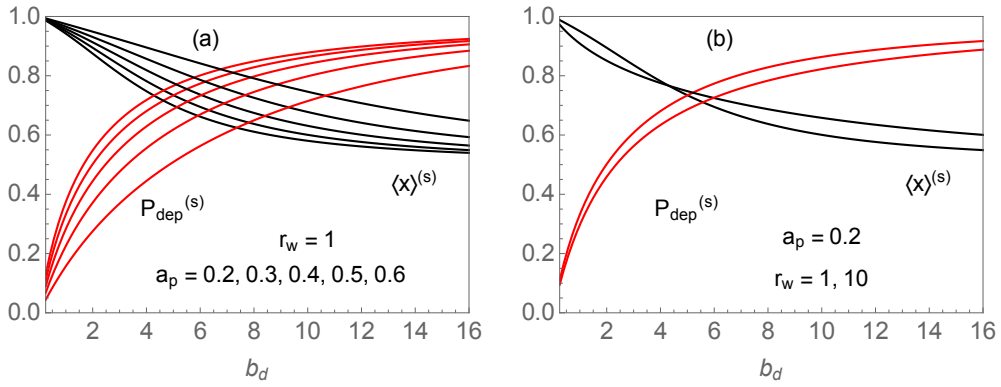

Figure 3.13. Dependence on parameters (a) $a_{\rm p}$, (b) $r_{\rm w}$ of mean scaled position $\langle x\rangle^{(\rm s)}$ and probability $P^{(\rm s)}_{\rm dep}$ versus parameter $0.25 \leq b_{\rm d} \leq 16$. With increasing $a_{\rm p}$ $\langle x\rangle^{(\rm s)}$ goes up and $P^{(\rm s)}_{\rm dep}$ goes down. With increasing $r_{\rm w}$ the trends (at large $b_{\rm d}$) are the same.

stationary states. The trajectory in time from any initial state toward the unique stationary state is expected by much more strongly affected by variations in $r_{\rm w}$.

3.5.2 Kinetics

We begin our investigation of the kinetics of insertion of the protonatable residue by tracking its motion in time for a given initial state under different circumstances. In Fig. 3.14 we show the position of the residue as a function of time under variations of two parameters.

Panel (a) represents a case of sluggish motion, $r_{\rm w} = 1$. The residue starts in the exterior region moves to the interface at high pH or to the center of the interior region at low pH. The residue gets to its destination on very similar time scales. It makes almost no difference wheter the inital state of the residue is protonated or deprotonated. We know from Fig. 3.11 that the final state is with high probability deprotonated at high pH and protonated at low pH.

The motion represented in panel (b) is significantly less sluggish compared to the pro/dep kinetics.: $r_{\rm w} = 10$. The physical time scale for the motion to be completed remains very similar. We continue using the time unit $c_{\rm f}^{-1}$. At high pH

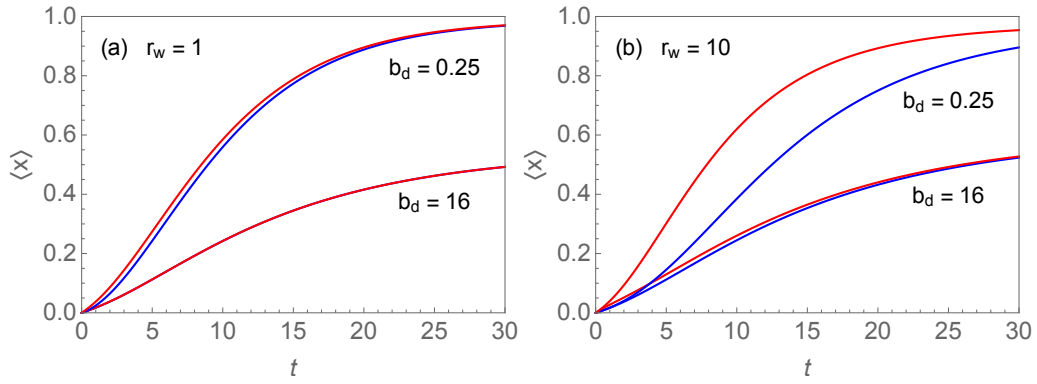

Figure 3.14. Average position $\langle x \rangle$ vs time t of protonatable residue launched in the exterior region in either a protonated state (red curve) or a deprotonated state (blue curve). Panels (a) and (b) use different relative time scales for the two intertwined processes. Each panel shows a case of low pH and a case of high pH. The parameter related to the change in pKa is kept at the value $a_{\rm p} = 0.5$ in all cases shown.

the initial potonation status has little impact on the average position. However, at low pH, the residue moves faster (on average) if it is initially protonated than if it is initially deprotonated.

This makes sense if we take into account that the force in the exterior region is stronger for the protonated residue than for the deprotonated residue. At high pH the (average) protonation goes down or stays down whereas at low pH it goes up or stays up.

Next we examine the kinetics of a situation where the protonatable residue is in the stationary state at high pH and then experiences a sudden drop in pH. The initial state is the one identified in Sec. 3.5.1 and described in the context of Fig. 3.11(a). The residue is, with high probability, deprotonated and located at the center of the interface.

When parameter controlling the pH is switched from the initial value, $b_{\rm d}^{\rm (ini)} = 16$ to one of four distinct actual values $b_{\rm d}^{\rm (act)} = 16, 8, 4, 0.25$ both the average protonation status and the average location change toward different values. The

long-time asymptotic states are the four stationary states described in the context of Fig. 3.11.

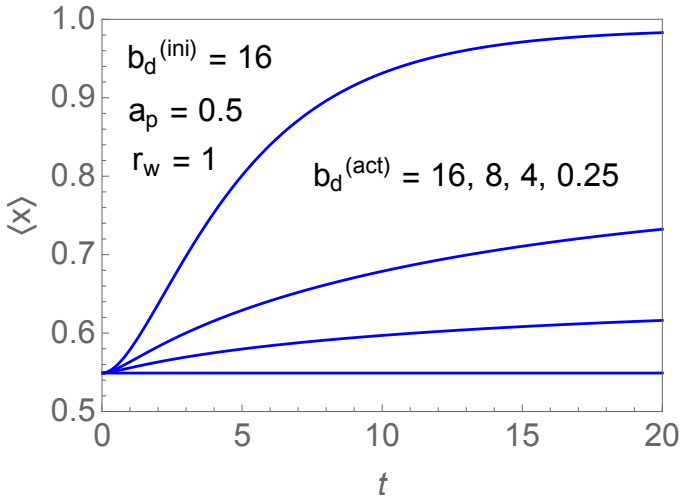

Figure 3.15. Mean scaled position $\langle x \rangle$ versus time of protonatable due to sudden change in pH. The initial state is the stationary state depicted in Fig. 3.11(a) at high pH. The long-time asymptotic state for each curve is the one of the four stationary states depicted n Fig. 3.11(a)-(d).

The time-dependence of the average position for the different cases are shown in Fig. 3.15. The final values of average position are consistent with the results shown in Fig. 3.12. Only for the largest change in the parameter value chosen does the average position move from the center of the interface to the center of the membrane. The shape of the curves in Fig. 3.15 tell us that the the kinetic process between initial state and (final) stationary state is faster at low pH than at intermediate pH. At $t = 20$ the slope of the top curve, pertaining to $b_{\rm d}^{(\rm act)} = 0.25$, is smaller that the slope of the next lower curve, pertaining to $b_{\rm d}^{(\rm act)} = 4$.

The observation that the average position increases gradually as we lower the value of the actual pH gradually does not mean that the most likely position of the residue changes gradually from near the center of the interface to near the center of the membrane. This fact is already evident in the stationary probability

distributions shown in Fig. 3.11. More details about the combined kinetics of relocation and protonation status is contained in Fig. 3.16.

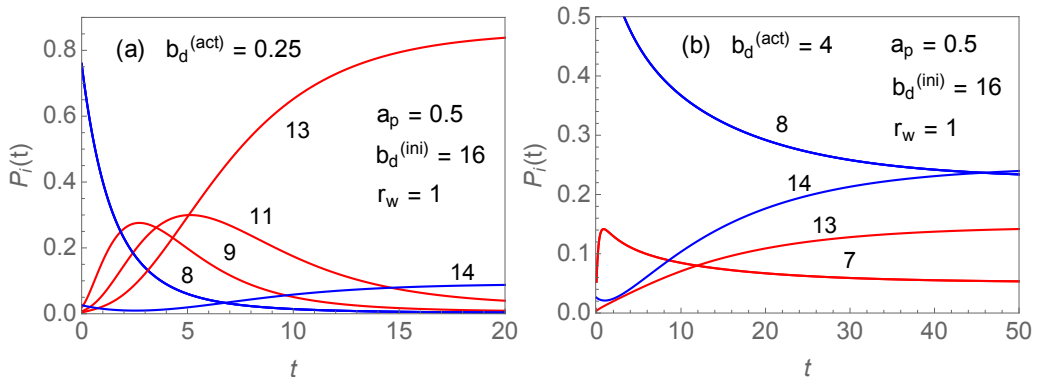

Figure 3.16. Probability $P_i(t)$ versus t of protonatable residue in state i. The initial state is the stationary state depicted in Fig. 3.11(a) at high pH. The long-time asymptotic state in panels (a) and (b) are the stationary states at low pH depicted in Fig. 3.11(d) and at intermediate pH depicted in Fig. 3.11(c), respectively. We only show curves for selected odd i (protonated states) and even i (deprotonated states). The curves not shown have lower maximum values. Note the different scales.

Consider first panel (a) pertaining to a large drop in pH. The residue starts out with high probability in state $i = 8$, which positions it at the center of the residue and has it deprotonated. As time evolves observe a quick change toward protonation and a slower change in location toward the center of the membrane. The probabilities at intermediate positions peak successively but the peak values are much higher for the protonated state (shown) than for the deprotonated state (not shown). Within 20 time units the kinetic process is very close to the long-time asymptotic state, which we already know from Sec. 3.5.1 as the stationary state at given pH. The growth with time of probability P_{14}, which specifies a deprotonated state, is simply a consequence of the fact that probability for the residue to be at that location increases. It does not indicate any reversal in protonation status.

Now consider panel (b) pertaining to an intermediate drop in pH. Here the

time scale related to kinetics of relocation is even slower compared to that of the kinetics of protonation/deprotonation. Note that we have extended the maximum time to reach closer to the long-time asymptotic state. One quick change in protonation status while the residue is still near the center of the interface is encoded in the sharp rise of probability P_9. All the probabilities shown then change more slowly toward their long-time asymptotic values. In this slower process we see the kinetics of relocation in action. The highest probabilities position the residue either near the center of the interface (states 7,8) or at the center of the membrane (states 13,14). The probabilities for the residue to be at intermediate positions are smaller. If the residue is still at the interface it is more likely to be deprotonated, if it has moved to the interior it is more likely to be protonated.

3.5.3 Variation of $\mathrm{pK_a}$

There is considerable uncertainty regarding the the amount and the rate by which the $\mathrm{pK_a}$ of the protonatable residues change between the (exterior) aqueous environment and the (interior) highly non-polar environment. While the modeling carried out thus far ties the rate of change to the amount of change, we can vary the amount through the model parameter a_p.

In Fig. 3.17 we show what the effect on the stationary probability distribution as discussed in the context of Fig. 3.11 is if we assume a significantly smaller or a significantly larger change in $\mathrm{pK_a}$. The larger change is represented by the value $a_\mathrm{p} = 0.2$ and the data point connected by solid lines whereas the smaller change is represented by the value $a_\mathrm{p} = 0.8$ and the data point connected by dashed lines. The case shown in Fig. 3.11 represents an intermediate change in $\mathrm{pK_a}$.

The main insight from this exercise is that that the insertion tendency is enhanced by a larger change in $\mathrm{pK_a}$. The difference between the two sets of data is highest at intermediate pH. In panel (c), for example, a small change in $\mathrm{pK_a}$

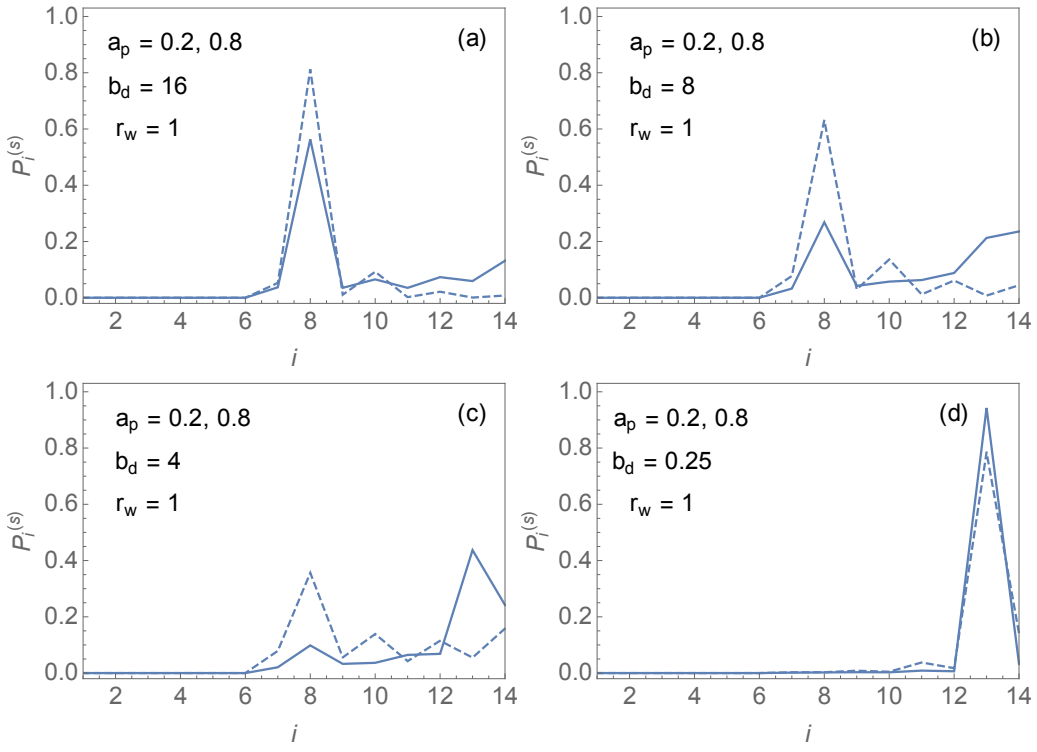

Figure 3.17. Stationary probabilities $P_i^{(\mathrm{s})}$, $i = 1, \ldots, 2N$ for $N = 7$. The four panels show data for different values of b_d, which controls the level of pH. The data in each panel pertain to two values of a_p, which controls the amount by which the $\mathrm{pK_a}$ changes between exterior and interior regions. The value $a_\mathrm{p} = 0.2$ reflects a large change (solid lines) and the value $a_\mathrm{p} = 0.8$ a small change.

has the residue still with high probability at the interface in a deprotonated state, whereas the a large change in $\mathrm{pK_a}$ has it with high probability at the center of the membrane in a protonated state. The enhanced tendency of insertion is already present at high pH as seen in panel (a). Only at low pH is the role of the parameter a_p of marginal significance.

The main insight already gained from stationarity is largely confirmed by results portraying the kinetics such as shown in Fig. 3.18. Comparison with the results of Fig. 3.15 for an intermediate change in $\mathrm{pK_a}$ we see that a large change in $\mathrm{pK_a}$ [panel (a)] enhances the trend toward insertion most strongly for intermediate values of pH. The exact opposite trend is exhibited by the data for a small change

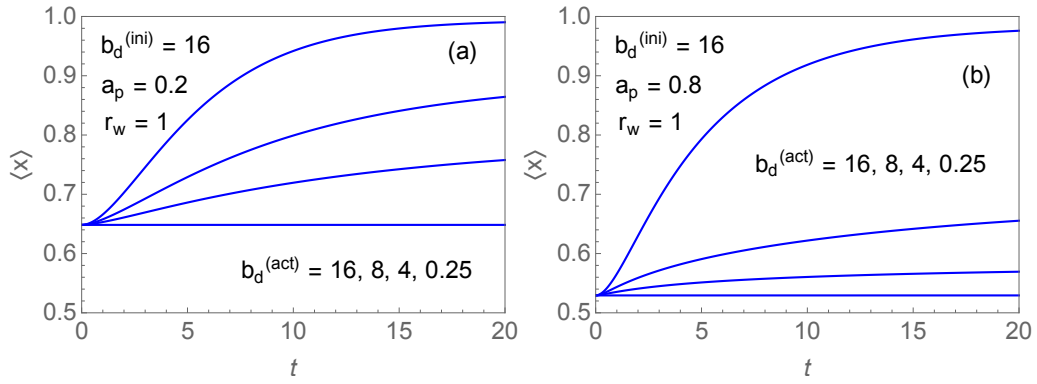

Figure 3.18. Mean scaled position $\langle x \rangle$ versus time of protonatable due to sudden change in pH. The initial state is the stationary state depicted in Fig. 3.17(a) at high pH. The long-time asymptotic state for each curve is the one of the four stationary states depicted in Fig. 3.17(a)-(d). Panel (a) for $a_{\rm p} = 0.2$ connects with the solid lines and panel (b) for $a_{\rm p} = 0.8$ connects with the dashed lines in Fig. 3.17.

in $\mathrm{pK_a}$ [panel (b)]. The trend towards insertion is equally strong for the case oh low pH despite the different (average) starting position.

3.5.4 Average-force approximation

The type of motion we have been considering throughout this study is over-damped and slow. However, what matters in our kinetic modeling is relative slowness, a measure captured by the parameter $r_{\rm w}$. It compares the speed of the pro/dep process with the speed of translocation. For $r_{\rm w} \simeq 1$ and smaller, the residue does not move significantly between pro/dep events whereas for $r_{\rm w} \gg 1$ it does

We recall that the pro/dep transition rates vary a great deal between the aqueous environment and the membrane interior. Their slowing down is attributable to two factors: the scarcity of free protons as reaction partners and the highly endothermic nature of proton production in the non-polar medium. Allowing for situations with large $r_{\rm w}$ has made the modeling significantly more complex than it

would be if $r_{\mathrm{w}} \lesssim 1$ could be guaranteed for all applications likely to be encountered.

Here we explore the limiting case in our modeling where the distance traveled between pro/dep events can be assumed to be tiny during the entire process. In this limit, the Markov chain model can be simplified as follows. Instead of the transition matrix $\mathbf{W}$, which is the product of the two non-commuting matrices $\mathbf{A}$ and $\mathbf{B}$, representing pro/dep status change and translocation, respectively, we now use a matrix $\mathbf{C}$ with the same structure as matrix $\mathbf{B}$ but modified elements.

The nonzero off-diagonal elements of $\mathbf{B}$ depend on the force experienced by the residue in a given pro/dep status as encoded in (3.12).

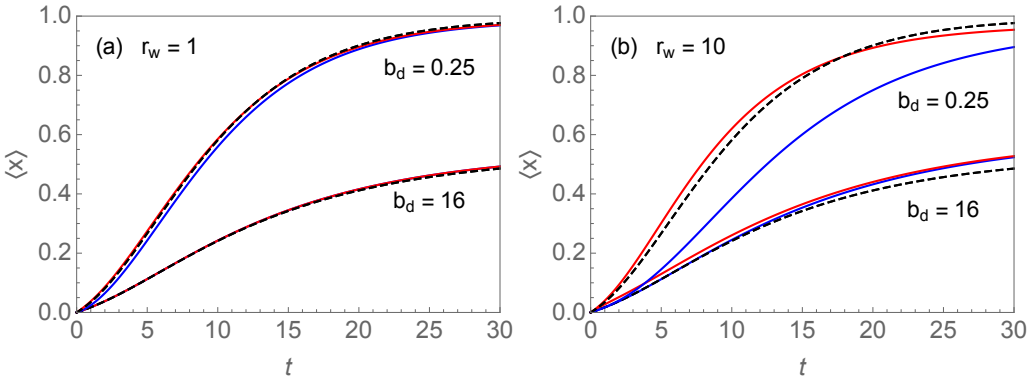

Figure 3.19. Average-force approximation data (dashed curves) added to the data alrady presented in Fig. 3.14.

The corresponding elements in the matrix $\mathbf{C}$, by contrast, depend on the average pro/dep status of the residue at any given position, as determined in Sec. 3.3.4. Keeping the same numbering of states, which now contains some redundancy, we write

$$C_{\mathrm{o,o\pm2}} = C_{\mathrm{e,e\pm2}} = P_{\mathrm{s}}^{(\mathrm{o})} B_{\mathrm{o,o\pm2}} + P_{\mathrm{s}}^{(\mathrm{e})} B_{\mathrm{e,e\pm2}} \tag{3.20}$$

for the off-diagonal elements and infer the diagonal elements from the normalization condition as in (3.12).

In Fig. 3.19 we show the data of Fig. 3.14 amended by data (shown dashed)

representing the average- force approximation. As expected the approximation is very accurate for $r_{\mathrm{w}} = 1$, where the pro/dep status equilibrates faster than the translocation, but turns unreliable at $r_{\mathrm{w}} = 10$, where significant motion between pro/dep are likely in some cases.

3.6 Project extensions

Naturally, the kinetics of pHLIP during its insertion into a membrane is more complex on several counts. It contains several protonatable residues and has a protonatable C terminus, all at different locations in the membrane environment and thus subject to different pro/dep transition rates. The relative positions between these residues and the C terminus is constrained in ways that depend on the conformation of the peptide (flexible coil or rigid α-helix).

In addition, we have to take into account, the forces acting on the other residues. The strongest inward forces will be experienced by the hydrophobic residues (e.g Leu) and the strongest outward forces by the positively charged Arg residue and N terminus. The relative locations of these sites are again constrained by the orientation and conformation of the peptide.

Here we have arrived at a point of closest contact with part two of the three-part modeling outlined in Sec. 3.2. In Ref. [17] we have explored pHLIP insertion pathways that are all downhill in the free-energy landscape or have to overcome, at most, small energy barriers. Free-energy landscapes are suggestive of realistic insertion pathways. They merely hint at the distinction between fast and slow pathways by the presence and size of energy barriers.

Insertion time scales of predictive significance require a kinetic study or a simulation study based on realistic rates for at least the dominant microscopic processes involved. This certainly includes the pro/dep transition rates and the rates of translocation of residues as investigated in Secs. 3.3-3.5 of this work. One

challenge for simulations is to deal with processes that involve a wide range of time scales. The approach taken in this study faces several challenges including the following.

The parameters $K_{\mathrm{p}}, a_{\mathrm{p}}, b_{\mathrm{d}}$, which have served the purpose for which they have been introduced reasonably well, must be more closely linked to chemical data such as pH in water, the $\mathrm{pK_w}$ of $\mathrm{H_2O}$ molecules in the non-polar lipid environment, and the $\mathrm{pK_a}$ of protonatable residues. The forces acting on individual residues in the modeling carried out thus far have strongly simplified profiles with a single parameter c_{f} setting the time scale for translocation. A significant refinement will be necessary in this aspect as well.

The design of our modeling is such that it is capable of accommodating these desired extensions. The number of discrete locations can readily be scaled up to improve spatial resolution. It is also straightforward to extend the space to include both sides of the membrane including a layer of water on either side at either the same or different levels of pH.

Our scheme of combining processes with individual time scales via controlled products of distinct transition matrices can be extended to include any number of relevant factors that govern the kinetics of insertion. As the level of complexity increases, a tensor notation of the Markov chain process will become more natural.

List of References

[1] R. E. Jacobs and S. H. White, "The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices," *Biochemistry*, vol. 28, no. 8, pp. 3421–3437, 1989.

[2] D. Engelman and T. Steitz, "The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis," *Cell*, vol. 23, no. 2, pp. 411–422, 1981.

[3] J.-L. Popot and D. M. Engelman, "Membrane protein folding and oligomerization: the two-stage model," *Biochemistry*, vol. 29, no. 17, pp. 4031–4037,

1990.

[4] S. H. White and W. C. Wimley, "Membrane protein folding and stability: physical principles," *Annual Review of Biophysics and Biomolecular Structure*, vol. 28, no. 1, pp. 319–365, 1999.

[5] M. Milik and J. Skolnick, "Insertion of peptide chains into lipid membranes: An off-lattice monte carlo dynamics model," *Proteins: Structure, Function, and Bioinformatics*, vol. 15, no. 1, pp. 10–25, 1993.

[6] A. S. Ladokhin and S. H. White, "Interfacial folding and membrane insertion of a designed helical peptide," *Biochemistry*, vol. 43, no. 19, pp. 5782–5791, 2004.

[7] M. J. Tucker, J. Tang, and F. Gai, "Probing the kinetics of membrane-mediated helix folding," *The Journal of Physical Chemistry B*, vol. 110, no. 15, pp. 8105–8109, 2006.

[8] J. Tang, R. S. Signarvic, W. F. DeGrado, and F. Gai, "Role of helix nucleation in the kinetics of binding of mastoparan x to phospholipid bilayers," *Biochemistry*, vol. 46, no. 48, pp. 13 856–13 863, 2007.

[9] M. Milik and J. Skolnick, "Spontaneous insertion of polypeptide chains into membranes: a monte carlo model." *Proceedings of the National Academy of Sciences*, vol. 89, no. 20, pp. 9391–9395, 1992.

[10] W. Im and C. L. Brooks, "Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 19, pp. 6771–6776, 2005.

[11] H. Nymeyer, T. B. Woolf, and A. E. Garcia, "Folding is not required for bilayer insertion: Replica exchange simulations of an α-helical peptide with an explicit lipid bilayer," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 4, pp. 783–790, 2005.

[12] M. B. Ulmschneider and J. P. Ulmschneider, "Membrane adsorption, folding, insertion and translocation of synthetic trans-membrane peptides," *Molecular Membrane Biology*, vol. 25, no. 3, pp. 245–257, 2008.

[13] M. B. Ulmschneider, J. C. Smith, and J. P. Ulmschneider, "Peptide partitioning properties from direct insertion studies," *Biophysical Journal*, vol. 98, no. 12, pp. L60–L62, 2010.

[14] M. B. Ulmschneider, J. P. Doux, J. A. Killian, J. C. Smith, and J. P. Ulmschneider, "Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides," *Journal of the American Chemical Society*, vol. 132, no. 10, pp. 3452–3460, 2010.

[15] Y. Deng, Z. Qian, Y. Luo, Y. Zhang, Y. Mu, and G. Wei, "Membrane binding and insertion of a phlip peptide studied by all-atom molecular dynamics simulations," *International Journal of Molecular Sciences*, vol. 14, no. 7, pp. 14 532–14 549, 2013.

[16] G. P. Sharma, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, "Coil-helix transition of polypeptide at water-lipid interface," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, p. P01034, 2015.

[17] G. P. Sharma, A. C. Meyer, Y. K. Reshetnyak, O. A. Andreev, M. Karbach, and G. Müller, "Free-energy landscapes for peptides in membrane environment," unpublished.

[18] Y. K. Reshetnyak, M. Segala, O. A. Andreev, and D. M. Engelman, "A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers," *Biophysical Journal*, vol. 93, no. 7, pp. 2363–2372, 2007.

[19] Y. K. Reshetnyak, O. A. Andreev, M. Segala, V. S. Markin, and D. M. Engelman, "Energetics of peptide (phlip) binding to and folding across a lipid bilayer membrane," *Proceedings of the National Academy of Sciences*, vol. 105, no. 40, pp. 15 340–15 345, 2008.

[20] D. Weerakkody, A. Moshnikova, M. S. Thakur, V. Moshnikova, J. Daniels, D. M. Engelman, O. A. Andreev, and Y. K. Reshetnyak, "Family of ph (low) insertion peptides for tumor targeting," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5834–5839, 2013.

[21] O. A. Andreev, A. D. Dupuy, M. Segala, S. Sandugu, D. A. Serra, C. O. Chichester, D. M. Engelman, and Y. K. Reshetnyak, "Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo," *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7893–7898, 2007.

[22] M. Musial-Siwek, A. Karabadzhak, O. A. Andreev, Y. K. Reshetnyak, and D. M. Engelman, "Tuning the insertion properties of phlip," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1798, no. 6, pp. 1041–1046, 2010.

[23] T. K. Harris and G. J. Turner, "Structural basis of perturbed pka values of catalytic groups in enzyme active sites," *IUBMB Life*, vol. 53, no. 2, pp. 85–98, 2002.

[24] F. N. Barrera, D. Weerakkody, M. Anderson, O. A. Andreev, Y. K. Reshetnyak, and D. M. Engelman, "Roles of carboxyl groups in the transmembrane insertion of peptides," *Journal of Molecular Biology*, vol. 413, no. 2, pp. 359–371, 2011.

[25] M. Zoonens, Y. K. Reshetnyak, and D. M. Engelman, "Bilayer interactions of phlip, a peptide that can deliver drugs and target tumors," *Biophysical Journal*, vol. 95, no. 1, pp. 225–235, 2008.

[26] F. N. Barrera, J. Fendos, and D. M. Engelman, "Membrane physical properties influence transmembrane helix formation," *Proceedings of the National Academy of Sciences*, vol. 109, no. 36, pp. 14 422–14 427, 2012.

[27] O. A. Andreev, A. G. Karabadzhak, D. Weerakkody, G. O. Andreev, D. M. Engelman, and Y. K. Reshetnyak, "ph (low) insertion peptide (phlip) inserts across a lipid bilayer as a helix and exits by a different path," *Proceedings of the National Academy of Sciences*, vol. 107, no. 9, pp. 4081–4086, 2010.

[28] A. G. Karabadzhak, D. Weerakkody, D. Wijesinghe, M. S. Thakur, D. M. Engelman, O. A. Andreev, V. S. Markin, and Y. K. Reshetnyak, "Modulation of the phlip transmembrane helix insertion pathway," *Biophysical Journal*, vol. 102, no. 8, pp. 1846–1855, 2012.

APPENDIX A

Polynomial equations and particle population densities

The 2μ Eqs. (1.6) with the $\beta\epsilon_m$ expressed via parameters (1.8) and (1.9) and the $g_{mm'}$ as stated in Sec. 1.2.2 acquire the following form (for $\mu \geq 2$):

$$\frac{t}{\tau} = \frac{w_1^2}{1+w_1}\frac{1+w_2}{w_2}\frac{1+w_{\mu+1}}{w_{\mu+1}}, \tag{A.1}$$

$$t^2 = (1+w_m)\frac{w_1^2}{(1+w_1)^2}\frac{1+w_{m+1}}{w_{m+1}}\frac{1+w_{\mu+m-1}}{w_{\mu+m-1}}\frac{1+w_{\mu+m}}{w_{\mu+m}}, \quad m = 2, ..., \mu-1 \tag{A.2}$$

$$t^2 = (1+w_\mu)\frac{w_1^2}{(1+w_1)^2}\frac{1+w_{2\mu-1}}{w_{2\mu-1}}\frac{1+w_{2\mu}}{w_{2\mu}}, \tag{A.3}$$

$$t = (1+w_m)\frac{w_1}{1+w_1}, \qquad m = \mu+1, ..., 2\mu. \tag{A.4}$$

From (A.4) we infer

$$w_{\mu+1} = ... = w_{2\mu} \doteq w, \tag{A.5}$$

which, upon substitution, simplifies (A.1)-(A.4) into

$$\frac{t}{\tau} = \frac{w_1^2}{1+w_1}\frac{1+w_2}{w_2}\frac{1+w}{w}, \tag{A.6}$$

$$t^2 = (1+w_m)\frac{w_1^2}{(1+w_1)^2}\frac{1+w_{m+1}}{w_{m+1}}\frac{(1+w)^2}{w^2}, \quad m = 2, ..., \mu-1, \tag{A.7}$$

$$t^2 = (1+w_\mu)\frac{w_1^2}{(1+w_1)^2}\frac{(1+w)^2}{w^2}, \tag{A.8}$$

$$t = (1+w)\frac{w_1}{1+w_1}. \tag{A.9}$$

Substitution of (A.9) into (A.6)-(A.8) yields

$$w_1 = \frac{w}{\tau}\frac{w_2}{1+w_2}, \quad \tau \neq 0, \tag{A.10}$$

$$w_m = w^2\frac{w_{m+1}}{1+w_{m+1}} - 1, \quad m = 2, ..., \mu-1, \tag{A.11}$$

$$w_\mu = w^2 - 1, \tag{A.12}$$

which express all w_m for $m \leq \mu$ from w recursively.

Next we show that these recursive relations can be satisfied by Chebyshev polynomials of the second kind, which themselves are generated recursively from $S_0(w) = 1$ and $S_1(w) = w$ via

$$S_{m+2}(w) = wS_{m+1}(w) - S_m(w), \quad m = 0, 1, 2, \ldots \tag{A.13}$$

We reason inductively by writing (A.12) in the form

$$w_m = \frac{S_{\mu-m+2}(w)}{S_{\mu-m}(w)}, \quad m = \mu. \tag{A.14}$$

If (A.14) holds for some m then we infer from (A.11) that it also holds for $m - 1$:

$$\begin{aligned} w_{m-1} &= w^2 \frac{w_m}{1 + w_m} - 1 = \frac{w^2 S_{\mu-m+2}(w)}{S_{\mu-m+2}(w) + S_{\mu-m}(w)} - 1 \\ &= \frac{wS_{\mu-m+2}(w) - S_{\mu-m+1}(w)}{S_{\mu-m+1}(w)} = \frac{S_{\mu-m+3}(w)}{S_{\mu-m+1}(w)}. \end{aligned} \tag{A.15}$$

This validates (A.14) for $m = 2, \ldots, \mu$. We use (A.10) to obtain

$$w_1 = \frac{S_\mu(w)}{\tau S_{\mu-1}(w)}. \tag{A.16}$$

The polynomial equation that determines w,

$$(1 + w - t)S_\mu(w) = t\tau S_{\mu-1}(w), \tag{A.17}$$

follows from (A.9) substituted in (A.16). This completes the derivation of (1.10) and (1.12). All w_m must be non-negative to be physically meaningful. Only one root of (A.17) satisfies this criterion.

The solution of the linear Eqs. (1.7) yields the following explicit expressions for the population densities of statistically interacting particles:

$$\begin{aligned} \bar{N}_m &= \frac{S_{\mu-m}(w)S_{\mu-m+1}(w)}{\gamma_\mu}, \\ \bar{N}_{\mu+m} &= \frac{[S_{\mu-m}(w)]^2}{\gamma_\mu}, \quad m = 1, \ldots, \mu, \end{aligned} \tag{A.18}$$

$$\gamma_\mu \doteq (1+w)\left[\sum_{m=0}^{\mu-1}\big[S_m(w)\big]^2 + \frac{\big[S_\mu(w)\big]^2}{t\tau}\right],$$
$$= (1+w)\left[\frac{2\mu+1-S_{2\mu}(w)}{4-w^2} + \frac{\big[S_\mu(w)\big]^2}{t\tau}\right]. \tag{A.19}$$

Entropy (1.14), enthalpy (1.15), helicity (1.16), and density of segments (1.17) can all be expressed in terms of the $\bar{N}_m$:

$$\frac{\bar{S}}{k_{\rm B}} = \sum_{m=1}^{2\mu}\bar{N}_m\Big[(1+w_m)\ln(1+w_m) - w_m\ln w_m\Big], \tag{A.20}$$

$$\bar{H} = \sum_{m=1}^{2\mu}\bar{N}_m\epsilon_m, \tag{A.21}$$

$$\bar{N}_{\rm seg} = \bar{N}_1 = \frac{\big[S_\mu(w)\big]^2}{\gamma_\mu}\frac{w+1-t}{t\tau}, \tag{A.22}$$

$$\bar{N}_{\rm hl} = 1-\bar{N}_1-2\sum_{m=2}^{\mu}\bar{N}_m - \sum_{m=\mu+1}^{2\mu}\bar{N}_m = \frac{\big[S_\mu(w)\big]^2}{\gamma_\mu}\frac{w+1}{t\tau}, \tag{A.23}$$

The shortest proof of (A.18) uses its substitution into a scaled version of (1.7),

$$w_{m'}\bar{N}_{m'} = \sum_{m=1}^{\mu}\big(g_{m'm}\bar{N}_m + g_{m',m+\mu}\bar{N}_{m+\mu}\big) = \delta_{1m'}. \tag{A.24}$$

We perform this substitution in four batches: (i) for $m' = 2\mu$ we use $g_{2\mu,m} = -\delta_{\mu m}$; (ii) for $m' = \mu + m''$, $m'' = 1,\ldots,\mu-1$ we use $g_{\mu+m'',m} = -\delta_{m''m} - \delta_{m'',m-1}$; (iii) for $m' = 2,\ldots,\mu$ we use $g_{m'm} = -\delta_{m',m+1}$; and (iv) for $m' = 1$ we use $g_{1m} = 2$, $g_{1,\mu+m} = 1$, $m = 1,\ldots,\mu$.

In the first three batches (A.24) is shown to be satisfied by merely using (1.12) and (A.18):

$$\text{(i)}:\ \ w[S_0(w)]^2 - S_0(w)S_1(w) = 0, \tag{A.25}$$

$$\text{(ii)}:\ \ w[S_{\mu-m''}(w)]^2 - S_{\mu-m''}(w)S_{\mu-m''+1}(w) - S_{\mu-m''-1}(w)S_{\mu-m''}(w) = 0, \tag{A.26}$$

$$\text{(iii)}:\ \ \frac{S_{\mu-m'+2}(w)}{S_{\mu-m'}(w)}S_{\mu-m'}(w)S_{\mu-m+1}(w) - S_{\mu-m'+1}(w)S_{\mu-m+2}(w) = 0. \tag{A.27}$$

In the fourth batch, (A.24) reduces to the identity,

$$\text{(iv)}: \quad \frac{[S_\mu(w)]^2}{\tau} + 2\sum_{m=1}^{\mu} S_{m-1}(w)S_m(w) + \sum_{m=1}^{\mu}[S_{m-1}(w)]^2 = \gamma_\mu, \tag{A.28}$$

which is proven by also using (1.10).

The two sources of disorder identified in Sec. 1.4.1, namely the disorder in the sequence of coil/helix segments of diverse lengths and disorder within individual coil segments, are related to the population densities $\bar{N}_m$ of 2μ species of particles from three catgories (hosts, hybrids, and tags).

Hosts ($m = 1$) generate coil segments out of the helix pseudo-vacuum whereas hybrids ($m = 2, \ldots, \mu$) and tags ($m = \mu + 1, \ldots, 2\mu$) extend coil segments at the expense of helix segments. Thermally excited hosts at random locations along the polypeptide helix thus produce one source of disorder and germinate the other source of disorder via the thermal excitation of hybrids and tags nested inside.

Each coil segment, nucleated by exactly one host particle, forms a self-avoiding random walk assembled from hybrids and tags. The distribution of hybrids and tags inside a large coil segment as realized in the limit $\tau \to 0$ at $t < t_0$ and inferred from (A.18), reads

$$\begin{aligned} \bar{N}_m &= \frac{2\sin\big((m-1)\phi_0\big)\sin(m\phi_0)}{(\mu+1)t_0}, \\ \bar{N}_{\mu+m} &= \frac{2\sin^2(m\phi_0)}{(\mu+1)t_0}, \quad m = 1, \ldots, \mu, \end{aligned} \tag{A.29}$$

where $\phi_0 = \pi/(\mu + 1)$. In the limit $\mu \to \infty$, the distributions of both hybrids and tags acquire identical $\sin^2 x$ density profiles if we set $x = m/\mu$ for hybrids and $x = (\mu + m)/\mu$ for tags.

APPENDIX B

Statistically interacting polymer links

The origin in quantum many-body theory of the methodology used here and its adaptation to problems of current interest in classical statistical mechanics has already been presented from several different angles. In chapter 1 we worked out one particular application with mathematical rigor. The extensions presented in Secs. 2.3 and 2.4 are all built from that foundation.

Here we summarize those results from chapter 1 that are being used as the main building blocks for these extensions. The microscopic model for the coil-helix transition of a long polypeptide at a water-lipid interface solved in Chapter 1 has three parameters: the growth parameter t, the nucleation parameter τ (both continuous) and the (discrete) parameter μ that numbers the coil states available to each residue.

For our extensions we only use the cases $\mu = 2$ and $\mu = \infty$ at $\tau > 0$ and we only use selected results. We quote the relevant expressions in ways that are easy to trace back to their derivations in chapter 1.

The scaled Gibbs free energy in closed form reads

$$\frac{\bar{G}(t,\tau)}{k_{\mathrm{B}}T} = -\ln\left(\frac{1+w(t,\tau)}{t}\right), \tag{B.1}$$

where, for $\mu = 2$,

$$w(t,\tau) = \frac{1}{3}\left[x + 2\sqrt{x^2+3y}\cos\frac{\varphi}{3}\right], \tag{B.2}$$
$$\tan\varphi = \frac{\sqrt{27(4y^3+y^2x^2+18yx^2+4x^4-27x^2)}}{x(2x^2+9y-27)},$$

with $x \doteq t-1$, $y \doteq 1+t\tau$ and $0 \le \varphi < \pi$, and, for $\mu = \infty$,

$$w(t,\tau) = \begin{cases} 2 & : 0 \le t \le t_{\mathrm{c}} \\ t-1+\dfrac{t\tau}{\lambda} & : t > t_{\mathrm{c}} \end{cases} \tag{B.3}$$

with

$$\lambda(t,\tau) \doteq \frac{1}{2}\left[t-1+\sqrt{(t+1)(t-3)+4t\tau}\right] \tag{B.4}$$

and

$$t_{\mathrm{c}} \doteq \frac{3}{1+\tau}. \tag{B.5}$$

The expressions for helicity (order parameter) and entropy, inferred from first derivatives of $\bar{G}$ become

$$\bar{N}_{\mathrm{hl}}(t,\tau) = \frac{t}{1+w}\frac{\partial w}{\partial t}, \tag{B.6}$$

$$\frac{\bar{S}(t,\tau)}{k_{\mathrm{B}}} = \ln(1+w) - \frac{1}{1+w}\left[t\ln t\frac{\partial w}{\partial t} + \tau\ln\tau\frac{\partial w}{\partial \tau}\right], \tag{B.7}$$

respectively. They are too unwieldy for $\mu = 2$ to be reproduced here but fairly concise for $\mu = \infty$:

$$N_{\mathrm{hl}}(t,\tau) = \begin{cases} 0 & : 0 \le t < t_c \\ 1 - \dfrac{t\tau}{\lambda^2 - 1 + t\tau} & : t \ge t_c \end{cases} \tag{B.8}$$

$$\begin{aligned}\frac{\bar{S}(t,\tau)}{k_{\mathrm{B}}} &= \ln\left(1+\frac{\tau}{\lambda}\right) \\ &\quad + \frac{t\tau}{\lambda^2-1+t\tau}\left(\ln t - \frac{\lambda^2-1}{t(\lambda+\tau)}\ln\tau\right).\end{aligned} \tag{B.9}$$

Finally the scaled enthalpy,

$$\begin{aligned}\frac{\bar{H}(t,\tau)}{k_{\mathrm{B}}T} &= \frac{\bar{G}(t,\tau)}{k_{\mathrm{B}}T} + \frac{\bar{S}(t,\tau)}{k_{\mathrm{B}}} \\ &= \frac{t\tau}{\lambda^2-1+t\tau}\left(\ln t - \frac{\lambda^2-1}{t(\lambda+\tau)}\ln\tau\right),\end{aligned} \tag{B.10}$$

is of importance in this work.

APPENDIX C

Rate constants of protonation and deprotonation

Here we present some details about our reasoning that led to the model parameters used in Sec. 3.3. We consider (protonatable) Asp residues in a range of locations between pure water at given level of pH and the center of the membrane. The focus is on the rates $k_{\rm p}$ and $k_{\rm d}$ of protonation and deprotonation, respectively, in particular their dependence on location and pH. The reaction at hand is

$$\mathrm{A}^- + \mathrm{H}^+ \leftrightharpoons \mathrm{AH}. \tag{C.1}$$

If in some volume element at a given location there are N Asp residues with n of them in the protonated state then we describe the pro/dep process by the master equation,

$$\frac{\partial}{\partial t}P(n,t) = \sum_m \Big[W(n|m)P(m,t) - W(m|n)P(n,t)\Big], \tag{C.2}$$

with transition rates,

$$W(m|n) = k_{\rm d} n \delta_{m,n-1} + k_{\rm p}(N-n)\delta_{m,n+1}. \tag{C.3}$$

and rate constants $k_{\rm p}$, $k_{\rm d}$.

The equation of motion for the average number, $\langle n(t)\rangle$, of protonated residues,

$$\frac{d}{dt}\langle n\rangle = \langle \alpha_1(n)\rangle, \tag{C.4}$$

depends on the first jump moment,

$$\alpha_1(n) \doteq \sum_m (m-n)W(m|n), \tag{C.5}$$

which, in this case, leads to the linear $1^{\rm st}$-order ODE,

$$\frac{d}{dt}\langle n(t)\rangle = k_{\rm p}N - k_+\langle n(t)\rangle, \tag{C.6}$$

where we define $k_+ \doteq k_\mathrm{d} + k_\mathrm{p}$. The solution for initial condition $\langle n(0)\rangle = n_0$ reads

$$\langle n(t)\rangle = n_0 e^{-k_+ t} + \frac{k_\mathrm{p} N}{k_+}\big(1 - e^{-k_+ t}\big). \tag{C.7}$$

The stationary solution of (C.6),

$$\frac{\langle n\rangle}{N} = \frac{1}{1 + k_\mathrm{d}/k_\mathrm{p}} = P_\mathrm{s}, \tag{C.8}$$

represents the probability P_s that any given Asp residue at that location is protonated. It is approached by (C.7) exponentially fast at the rate k_+.

The ratio $k_\mathrm{d}/k_\mathrm{p}$ is determined by the dissociation equilibrium. In pure water we have

$$\big[\mathrm{H^+}\big] = 10^{-\mathrm{pH}}, \quad \frac{\big[\mathrm{A^-}\big]\big[\mathrm{H^+}\big]}{\big[\mathrm{AH}\big]} = 10^{-\mathrm{pK_a}}, \tag{C.9}$$

and, in any environment,

$$P_\mathrm{s} = \frac{\big[\mathrm{AH}\big]}{\big[\mathrm{AH}\big] + \big[\mathrm{A^-}\big]} = \frac{1}{1 + \big[\mathrm{A^-}\big]/\big[\mathrm{AH}\big]}. \tag{C.10}$$

Comparison of (C.10) with (C.8) with use of (C.9) yields the key result (3.1) used in Sec. 3.3:

$$\frac{k_\mathrm{d}}{k_\mathrm{p}} = 10^{\mathrm{pH}-\mathrm{pK_a}}. \tag{C.11}$$

Given that $\mathrm{pK_a} \simeq 4$ for Asp in water, the ratio (C.11) is large at neutral pH and becomes small at sufficiently low pH, implying that Asp is very likely to be deprotonated at neutral pH and to become protonated when the aqueous environment turns sufficiently acidic.

At locations away from pure water toward the membrane interior, the dissociation equilibrium condition is more complex. It now depends on the $\mathrm{pK_w}$ of $\mathrm{H_2O}$ in addition to the $\mathrm{pK_a}$ of Asp. Both quantities are expected to rise. Evidence from MD simulations [Vila-Vicosa et al. 2018] strongly suggests that the $\mathrm{pK_a}$ of Asp rises from 4 to near 7 between water and membrane interior. The $\mathrm{pK_w}$ likewise

rises from twice the pH to a value yet to be estimated. It is reasonable to expect that if the pH in water is lowered, then the pK_w will go down throughout the membrane environment.

A different chain of reasoning can be employed to estimate the dependence on position of k_d and k_p in the membrane environment. The relative dielectric constant, $\epsilon(x)$ is known to vary between a high value, $\epsilon_1 \simeq 80$, in pure water and a much lower value, $\epsilon_2 \simeq 5$, inside the membrane. Key for our argument is that all electrostatic energies are reduced by a factor $\epsilon(x)^{-1}$.

Deprotonation creates two charges. Its energetic cost rises steeply between polar and non-polar environments. The activation energy per charge associated with deprotonation thus contains a factor $\epsilon(x)^{-1}$. It is then fair to conclude that the rate constant of deprotonation of a neutral Asp residue is suppressed as follows:

$$k_d \propto e^{-2\beta E_e/\epsilon(x)}. \tag{C.12}$$

The rate constant of protonation of a charged Asp residue at a given location x is also suppressed but less so and only indirectly due to the suppressed availability of protons. Only the charge of the proton affects the rate. The suppressed density of deprotonated residues itself is already accounted for by the factor $(N - n)$ in (C.3). Therefore, we can write

$$k_p \propto e^{-\beta E_e/\epsilon(x)}, \tag{C.13}$$

which, with (C.12), implies

$$\frac{k_d}{k_p} \propto e^{-\beta E_e/\epsilon(x)}. \tag{C.14}$$

If this reasoning is sound then it follows that both rate constants are exponentially suppressed on the way from water into the membrane but the ratio is also exponentially suppressed. Comparison of (C.14) with (C.11) then implies that the pK_w rises more slowly than the pK_a as the location of the protonatable residue

changes from water toward the center of the membrane. Our modeling of the matrix elements in (3.2) are based on these conclusions.

BIBLIOGRAPHY

Andreev, O. A., Dupuy, A. D., Segala, M., Sandugu, S., Serra, D. A., Chichester, C. O., Engelman, D. M., and Reshetnyak, Y. K., "Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo," *Proceedings of the National Academy of Sciences*, vol. 104, no. 19, pp. 7893–7898, 2007.

Andreev, O. A., Karabadzhak, A. G., Weerakkody, D., Andreev, G. O., Engelman, D. M., and Reshetnyak, Y. K., "ph (low) insertion peptide (phlip) inserts across a lipid bilayer as a helix and exits by a different path," *Proceedings of the National Academy of Sciences*, vol. 107, no. 9, pp. 4081–4086, 2010.

Anghel, D., "The fractional exclusion statistics amended," *EPL (Europhysics Letters)*, vol. 87, no. 6, p. 60009, 2009.

Bakhti, B., Maass, M., Karbach, M., and Müller, G., "Monodisperse rods in external potentials," *Phys. Rev. E*, vol. 92, p. 042112, 2015.

Bakhti, B., Karbach, M., Maass, P., Mokim, M., and Müller, G., "Statistically interacting vacancy particles," *Physical Review E*, vol. 89, no. 1, p. 012137, 2014.

Barrera, F. N., Fendos, J., and Engelman, D. M., "Membrane physical properties influence transmembrane helix formation," *Proceedings of the National Academy of Sciences*, vol. 109, no. 36, pp. 14 422–14 427, 2012.

Barrera, F. N., Weerakkody, D., Anderson, M., Andreev, O. A., Reshetnyak, Y. K., and Engelman, D. M., "Roles of carboxyl groups in the transmembrane insertion of peptides," *Journal of Molecular Biology*, vol. 413, no. 2, pp. 359–371, 2011.

Burkhardt, T. W., Yang, Y., and Gompper, G., "Fluctuations of a long, semiflexible polymer in a narrow channel," *Physical Review E*, vol. 82, no. 4, p. 041801, 2010.

Carlon, E., Orlandini, E., and Stella, A. L., "Roles of stiffness and excluded volume in dna denaturation," *Physical Review Letters*, vol. 88, no. 19, p. 198101, 2002.

Causo, M. S., Coluzzi, B., and Grassberger, P., "Simple model for the dna denaturation transition," *Physical Review E*, vol. 62, no. 3, p. 3958, 2000.

Chakrabarti, B. and Levine, A. J., "Nonlinear elasticity of an α-helical polypeptide," *Physical Review E*, vol. 71, no. 3, p. 031905, 2005.

Chakrabarti, B. and Levine, A. J., "Nonlinear elasticity of an α-helical polypeptide: Monte carlo studies," *Physical Review E*, vol. 74, no. 3, p. 031903, 2006.

Chen, Y., Zhang, Q., and Ding, J., "A coarse-grained model and associated lattice monte carlo simulation of the coil–helix transition of a homopolypeptide," *The Journal of Chemical Physics*, vol. 120, no. 7, pp. 3467–3474, 2004.

Cule, D. and Hwa, T., "Denaturation of heterogeneous dna," *Physical Review Letters*, vol. 79, no. 12, p. 2375, 1997.

Dauxios, T., Peyrard, M., and Bishop, A. R., "Entropy-driven dna denaturation," *Physical Review E*, vol. 47, no. 684, p. R44, 1993.

Deng, Y., Qian, Z., Luo, Y., Zhang, Y., Mu, Y., and Wei, G., "Membrane binding and insertion of a phlip peptide studied by all-atom molecular dynamics simulations," *International Journal of Molecular Sciences*, vol. 14, no. 7, pp. 14 532–14 549, 2013.

Einert, T. R. and Netz, R. R., "Theory for rna folding, stretching, and melting including loops and salt," *Biophysical Journal*, vol. 100, no. 11, pp. 2745–2753, 2011.

Engelman, D. and Steitz, T., "The spontaneous insertion of proteins into and across membranes: the helical hairpin hypothesis," *Cell*, vol. 23, no. 2, pp. 411–422, 1981.

Engelman, D. M., Chen, Y., Chin, C.-N., Curran, A. R., Dixon, A. M., Dupuy, A. D., Lee, A. S., Lehnert, U., Matthews, E. E., Reshetnyak, Y. K., *et al.*, "Membrane protein folding: beyond the two stage model," *FEBS Letters*, vol. 555, no. 1, pp. 122–125, 2003.

Finkelstein, A. V. and Ptitsyn, O., *Protein Physics*. AIP Press, New York, 2002.

Fisher, M. E., "Effect of excluded volume on phase transitions in biopolymers," *The Journal of Chemical Physics*, vol. 45, no. 5, pp. 1469–1473, 1966.

Grosberg, A. and Khokhlov, A., *Statistical Physics of Macromolecules*. AIP Press, New York, 1994.

Gundlach, N., Karbach, M., Liu, D., and Müller, G., "Jammed disks in a narrow channel: criticality and ordering tendencies," *J. Stat. Mech*, vol. 2013, no. 1, p. P04018, 2013.

Haldane, F. D. M., "Fractional statistics in arbitrary dimensions: A generalization of the pauli principle," *Physical Review Letters*, vol. 67, no. 8, p. 937, 1991.

Hanke, F., Serr, A., Kreuzer, H. J., and Netz, R. R., "Stretching single polypeptides: The effect of rotational constraints in the backbone," *EPL (Europhysics Letters)*, vol. 92, no. 5, p. 53001, 2010.

Harris, T. K. and Turner, G. J., "Structural basis of perturbed pka values of catalytic groups in enzyme active sites," *IUBMB Life*, vol. 53, no. 2, pp. 85–98, 2002.

Hofsäß, C., Lindahl, E., and Edholm, O., "Molecular dynamics simulations of phospholipid bilayers with cholesterol," *Biophysical Journal*, vol. 84, no. 4, pp. 2192–2206, 2003.

Im, W. and Brooks, C. L., "Interfacial folding and membrane insertion of designed peptides studied by molecular dynamics simulations," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 19, pp. 6771–6776, 2005.

Isakov, S. B., "Generalization of statistics for several species of identical particles," *Modern Physics Letters B*, vol. 8, no. 05, pp. 319–327, 1994.

Jacobs, R. E. and White, S. H., "The nature of the hydrophobic binding of small peptides at the bilayer interface: implications for the insertion of transbilayer helices," *Biochemistry*, vol. 28, no. 8, pp. 3421–3437, 1989.

Kafri, Y., Mukamel, D., and Peliti, L., "Why is the dna denaturation transition first order?" *Physical Review Letters*, vol. 85, no. 23, p. 4988, 2000.

Karabadzhak, A. G., Weerakkody, D., Wijesinghe, D., Thakur, M. S., Engelman, D. M., Andreev, O. A., Markin, V. S., and Reshetnyak, Y. K., "Modulation of the phlip transmembrane helix insertion pathway," *Biophysical Journal*, vol. 102, no. 8, pp. 1846–1855, 2012.

Ladokhin, A. S. and White, S. H., "Interfacial folding and membrane insertion of a designed helical peptide," *Biochemistry*, vol. 43, no. 19, pp. 5782–5791, 2004.

Lamura, A., Burkhardt, T., and Gompper, G., "Helical polymer in cylindrical confining geometries," *Physical Review E*, vol. 70, no. 5, p. 051804, 2004.

Lifson, S., "Partition functions of linear-chain molecules," *The Journal of Chemical Physics*, vol. 40, no. 12, pp. 3705–3710, 1964.

Liu, D., Lu, P., Müller, G., and Karbach, M., "Taxonomy of particles in ising spin chains," *Physical Review E*, vol. 84, no. 2, p. 021136, 2011.

Liu, D., Vanasse, J., Müller, G., and Karbach, M., "Generalized pauli principle for particles with distinguishable traits," *Physical Review E*, vol. 85, no. 1, p. 011144, 2012.

Lu, P., Liu, D., and Müller, "Interlinking motifs and entropy landscapes of statistically interacting particles," *Condensed Matter Physics*, vol. 15, p. 13001, 2012.

MacCallum, J. L., Bennett, W. D., and Tieleman, D. P., "Distribution of amino acids in a lipid bilayer from computer simulations," *Biophysical Journal*, vol. 94, no. 9, pp. 3393–3404, 2008.

Meyer, A., Gundlach, N., Karbach, M., Lu, P., and Müller, G., "Molecular chains under tension and torque: Force-extension and torque-twist characteristics," unpublished.

Milik, M. and Skolnick, J., "Spontaneous insertion of polypeptide chains into membranes: a monte carlo model." *Proceedings of the National Academy of Sciences*, vol. 89, no. 20, pp. 9391–9395, 1992.

Milik, M. and Skolnick, J., "Insertion of peptide chains into lipid membranes: An off-lattice monte carlo dynamics model," *Proteins: Structure, Function, and Bioinformatics*, vol. 15, no. 1, pp. 10–25, 1993.

Moore, C., Liu, D., Ballnus, B., Karbach, M., and Müller, G., "Disks in narrow channel jammed by gravity and centrifuge: profiles of pressure," *J. Stat. Mech*, vol. 2014, p. P04008, 2014.

Musial-Siwek, M., Karabadzhak, A., Andreev, O. A., Reshetnyak, Y. K., and Engelman, D. M., "Tuning the insertion properties of phlip," *Biochimica et Biophysica Acta (BBA)-Biomembranes*, vol. 1798, no. 6, pp. 1041–1046, 2010.

Nemnes, G. A. and Anghel, D.-V., "Fractional exclusion statistics in non-homogeneous interacting particle systems," *Roman. Rep. Phys.*, vol. 66, no. 336, 2014.

Nowak, C., Rostiashvili, V. G., and Vilgis, T. A., "Globular structures of a helix-coil copolymer: Self-consistent treatment," *The Journal of Chemical Physics*, vol. 126, no. 3, p. 034902, 2007.

Nymeyer, H., Woolf, T. B., and Garcia, A. E., "Folding is not required for bilayer insertion: Replica exchange simulations of an α-helical peptide with an explicit lipid bilayer," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 4, pp. 783–790, 2005.

Peyrard, M. and Bishop, A. R., "Statistical mechanics of a nonlinear model for dna denaturation," *Physical Review Letters*, vol. 62, no. 23, p. 2755, 1989.

Poland, D. and Scheraga, H. A., *Theory of helix-coil transitions in biopolymers: statistical mechanical theory of order-disorder transitions in biological macromolecules.* Academic Press, 1970.

Poland, D. and Scheraga, H. A., "Phase transitions in one dimension and the helix-coil transition in polyamino acids," *The Journal of Chemical Physics*, vol. 45, no. 5, pp. 1456–1463, 1966.

Popot, J.-L. and Engelman, D. M., "Membrane protein folding and oligomerization: the two-stage model," *Biochemistry*, vol. 29, no. 17, pp. 4031–4037, 1990.

Potter, G. G., Müller, G., and Karbach, M., "Thermodynamics of ideal quantum gas with fractional statistics in d dimensions," *Phys. Rev. E*, vol. 75, p. 061112, 2007.

Ramachandran, S., Komura, S., Seki, K., and Gompper, G., "Dynamics of a polymer chain confined in a membrane," *The European Physical Journal E*, vol. 34, no. 5, p. 46, 2011.

Reshetnyak, Y. K., Andreev, O. A., Segala, M., Markin, V. S., and Engelman, D. M., "Energetics of peptide (phlip) binding to and folding across a lipid bilayer membrane," *Proceedings of the National Academy of Sciences*, vol. 105, no. 40, pp. 15 340–15 345, 2008.

Reshetnyak, Y. K., Segala, M., Andreev, O. A., and Engelman, D. M., "A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers," *Biophysical Journal*, vol. 93, no. 7, pp. 2363–2372, 2007.

Sharma, G. P., Meyer, A. C., Reshetnyak, Y. K., Andreev, O. A., Karbach, M., and Müller, G., "Free-energy landscapes for peptides in membrane environment," unpublished.

Sharma, G. P., Reshetnyak, Y. K., Andreev, O. A., Karbach, M., and Müller, G., "Coil-helix transition of polypeptide at water-lipid interface," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, p. P01034, 2015.

Stepanov, V. V., Müller, G., and Stolze, J., "Quantum integrability and nonintegrability in the spin-boson model," *Phys. Rev. E*, vol. 77, p. 066202, 2008.

Tamashiro, M. N. and Pincus, P., "Helix-coil transition in homopolypeptides under stretching," *Physical Review E*, vol. 63, no. 2, p. 021909, 2001.

Tang, J., Signarvic, R. S., DeGrado, W. F., and Gai, F., "Role of helix nucleation in the kinetics of binding of mastoparan x to phospholipid bilayers," *Biochemistry*, vol. 46, no. 48, pp. 13 856–13 863, 2007.

Theodorakopoulos, N., Dauxois, T., and Peyrard, M., "Order of the phase transition in models of dna thermal denaturation," *Physical Review Letters*, vol. 85, no. 1, p. 6, 2000.

Tucker, M. J., Tang, J., and Gai, F., "Probing the kinetics of membrane-mediated helix folding," *The Journal of Physical Chemistry B*, vol. 110, no. 15, pp. 8105–8109, 2006.

Ulmschneider, M. B., Doux, J. P., Killian, J. A., Smith, J. C., and Ulmschneider, J. P., "Mechanism and kinetics of peptide partitioning into membranes from all-atom simulations of thermostable peptides," *Journal of the American Chemical Society*, vol. 132, no. 10, pp. 3452–3460, 2010.

Ulmschneider, M. B., Smith, J. C., and Ulmschneider, J. P., "Peptide partitioning properties from direct insertion studies," *Biophysical Journal*, vol. 98, no. 12, pp. L60–L62, 2010.

Ulmschneider, M. B. and Ulmschneider, J. P., "Membrane adsorption, folding, insertion and translocation of synthetic trans-membrane peptides," *Molecular Membrane Biology*, vol. 25, no. 3, pp. 245–257, 2008.

Vila-Viçosa, D., Silva, T. F., Slaybaugh, G., Reshetnyak, Y. K., Andreev, O. A., and Machuqueiro, M., "Membrane-induced pka shifts in wt-phlip and its l16h variant," *Journal of Chemical Theory and Computation*, vol. 14, no. 6, pp. 3289–3297, 2018.

Wang, L., "Measurements and implications of the membrane dipole potential," *Annual Review of Biochemistry*, vol. 81, pp. 615–635, 2012.

Weerakkody, D., Moshnikova, A., Thakur, M. S., Moshnikova, V., Daniels, J., Engelman, D. M., Andreev, O. A., and Reshetnyak, Y. K., "Family of ph (low) insertion peptides for tumor targeting," *Proceedings of the National Academy of Sciences*, vol. 110, no. 15, pp. 5834–5839, 2013.

White, S. H. and Wimley, W. C., "Membrane protein folding and stability: physical principles," *Annual Review of Biophysics and Biomolecular Structure*, vol. 28, no. 1, pp. 319–365, 1999.

Wu, Y.-S., "Statistical distribution for generalized ideal gas of fractional-statistics particles," *Physical Review Letters*, vol. 73, no. 7, p. 922, 1994.

Yang, Y., Burkhardt, T. W., and Gompper, G., "Free energy and extension of a semiflexible polymer in cylindrical confining geometries," *Physical Review E*, vol. 76, no. 1, p. 011804, 2007.

Zimm, B. H. and Bragg, J., "Theory of the phase transition between helix and random coil in polypeptide chains," *The journal of Chemical Physics*, vol. 31, no. 2, pp. 526–535, 1959.

Zoonens, M., Reshetnyak, Y. K., and Engelman, D. M., "Bilayer interactions of phlip, a peptide that can deliver drugs and target tumors," *Biophysical Journal*, vol. 95, no. 1, pp. 225–235, 2008.