Symmetry-Breaking Motility and RNA Secondary Structures

by

Allen Lee

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Master of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2005

Author .....
Department of Physics
August 4, 2005

Certified by .....
Mehran Kardar
Professor
Thesis Supervisor

Accepted by .....
THOMAS J. GREYTAK ~~Marc Kastner~~
Department of Physics

Symmetry-Breaking Motility and RNA Secondary Structures

by

Allen Lee

Submitted to the Department of Physics
on August 5, 2005, in partial fulfillment of the
requirements for the degree of
Master of Science in Physics

Abstract

This thesis contains work on three separate topics: the spontaneous motility of functionalized particles, the designability of RNA secondary structures, and the statistical mechanics of homopolymer RNAs.

For the work on spontaneous motility, we were motivated by *in vitro* experiments investigating the symmetry-breaking motility of functionalized spherical beads to develop a general theory for the dynamics of a rigid object propelled by an active process at its surface. Starting from a phenomenological expansion for the microscopic dynamics, we derive equations governing the macroscopic velocities of the object near an instability towards spontaneous motion. These equations respect symmetries in the object's shape, with implications for the phase behavior and singularities encountered at a continuous transition between stationary and moving states. Analysis of the velocity fluctuations of such an object reveals that these fluctuations differ qualitatively from those of a passive object.

For the work on designability, we investigated RNA folding within a toy model in which RNA bases come in two types and complementary base pairing is favored. Following a geometric formulation of biopolymer folding proposed in the literature, we represent RNA sequences and structures by points in a high-dimensional "contact space." Designability is probed by investigating the distribution of sequence and structure points within this space. We find that one-dimensional projections of the sequence point distribution approach normality with increasing RNA length N. Numerical comparison of the structure point distribution with a Gaussian approximation generated by principal component analysis reveals discrepancies.

The third and final project concerns the statistical mechanics of homopolymer RNAs. We compute the asymptotics of the partition function Z_n and characterize the crossover length scale governing its approach to its leading asymptotic behavior. Consideration of restricted partition functions in which one or more base pairs are enforced leads to an interesting connection with ideal Gaussian polymers. We introduce the notion of gapped secondary structures and analyze the partition function $Z_{n,p}^{(g)}$ for RNAs of length n with gap at p. Another length scale emerges whose scaling agrees with that of the crossover scale found earlier.

Thesis Supervisor: Mehran Kardar
Title: Professor

Acknowledgments

I thank my advisors Mehran Kardar and Alexander van Oudenaarden for their guidance and understanding, Elisa Gabbert for her companionship while most of this work was being done, and the National Science Foundation for financial support. I also take this opportunity to express my gratitude towards my family and country for all the advantages and comforts I enjoy in life.

Contents

1 Symmetry-Breaking Motility **9**

1.1 Introduction 9

1.2 Model 11

1.3 Analysis 13

1.3.1 Spherical Beads 13

1.3.2 Nonspherical Beads 15

1.3.3 Rotational Motion 18

1.3.4 Beads in External Potentials 22

1.4 Fluctuations 23

1.4.1 Spherical Beads 23

1.4.2 Non-Maxwellian Fluctuations 24

1.5 Conclusions and Discussion 26

2 RNA Secondary Structures: Designability and Combinatorics **31**

2.1 Introduction 31

2.2 RNA Folding and Designability 33

2.2.1 Introduction 33

2.2.2 Model 35

2.2.3 The Distribution of Sequence Points 36

2.2.4 The Distribution of Structure Points 39

2.2.5 Discussion 42

2.3 Gapped RNA Secondary Structures 43

2.3.1 The Homopolymer RNA Partition Function Z_n 43

2.3.2 The Restricted Partition Function $Z^{(r)}_{n,i,j}$ and Gapped Secondary Structures 46
2.3.3 Analysis of the Gapped RNA Partition Function $Z^{(g)}_{n,p}$ 48
2.3.4 Discussion 51

A Results on the Distribution of Sequence Points **53**

B Results on RNA Combinatorics **57**

Chapter 1

Symmetry-Breaking Motility

1.1 Introduction

The pathogenic bacteria *Listeria monocytogenes* has become a model system for investigations into actin-based motility. These bacteria are coated with a protein, ActA, that catalyzes actin polymerization in the cytoplasm of a host. After infection, this coating facilitates the spontaneous formation of a polarized actin "comet tail" that pushes *Listeria* forward, propelling it rapidly at rates up to over a micron per second [10]. Besides being an interesting biophysical phenomenon in its own right, it is hoped that understanding *Listeria's* actin-based propulsion will give insight into the cytoskeletal processes underlying eukaryotic cell motility and organization [6].

A number of *in vitro* experiments have been performed using artificial systems designed to mimic *Listeria* motility under simplified and more easily controlled conditions [35, 12, 33, 28, 5, 7]. In these experiments, the bacteria are replaced with spherical beads coated with ActA or a functionally similar protein such as N-WASP. When these functionalized beads are placed in cell extracts or a mixture of purified proteins, they can acquire *Listeria*-like motility, even if the coating of the beads is uniform. When such uniformly coated beads are initially placed in solution, they first become enveloped in a spherically symmetric cloud of actin filaments. Then, depending on the various parameters of the experiment, some beads are observed to spontaneously break free of their actin clouds and move away, propelled by an actin

comet tail. This symmetry-breaking phenomenon can be quite striking [1].

Several theories have been put forward to explain this spontaneous motility. van Oudenaarden and Theriot model the actin filaments surrounding the bead as elastic Brownian ratchets [32]. The stochastic growth of these filaments is coupled by the bead so that the growth of one filament makes room for the growth of neighboring filaments, generating positive feedback and the possibility of avalanching symmetry-breaking. Mogilner and Oster suggest an alternative explanation based on the cooperative failure of cross-links within the actin network [24]. Another type of model treats the actin cloud surrounding the bead as an elastic gel [25, 30, 13]. Polymerization occurs at the bead surface, expanding and stretching the gel until the resulting stress causes the gel's outer surface to rupture, allowing the bead to escape. The rupture can either be abrupt, as in the fracture of a brittle material [25, 13], or alternatively it may arise continuously from small fluctuations in the gel thickness [30].

Our aim is not to reconcile these existing theories, or to provide a comprehensive description of the observed phenomena. Instead, we investigate the behavior of an activated rigid bead within a framework that is sufficiently general to include many possible microscopic models. In the spirit of the Landau theory of magnetism, we treat the actin filaments as an effective field that generates a normal force density at each point on the bead surface. The time evolution of this field depends on the bead velocity, allowing the possibility of positive feedback as in the Brownian ratchet model [32]. Our approach neglects some aspects of the experimental system frequently mentioned in the literature, such as nonlocal elastic effects arising from the crosslinking of filaments [30]. In addition, as will be discussed later, a number of idealizations are made that may misrepresent the experimental situation(s) to varying degrees. While the theory has its weaknesses as far as the particular actin-based phenomena in consideration are concerned, its generality makes it relevant to a larger class of phenomena in which any rigid object is propelled by forces at its surface.

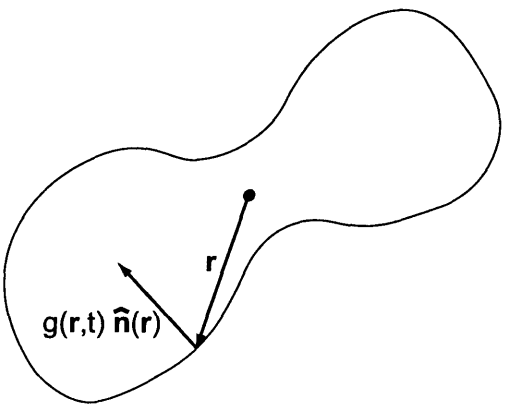

Figure 1-1: Rigid bead surrounded by a cloud of polymerizing actin filaments. The filaments exert a force per unit area $g(\mathbf{r}, t)$ on the surface of the bead.

1.2 Model

We consider a rigid bead with a given fixed shape. Points on the surface of the bead are specified by a vector $\mathbf{r}$ measured from the bead's center of mass. We denote the unit normal vector at $\mathbf{r}$ by $\hat{\mathbf{n}}(\mathbf{r})$. Our model for the dynamics of a functionalized bead consists of the following elements:

1. Actin polymerization (or any other active, energy-consuming process) exerts a force locally directed along the normal at every point on the bead surface. The force per unit area at time t and location $\mathbf{r}$ is denoted by $g(\mathbf{r}, t)$. See Fig. 1-1. The net force on the bead due to the activity is the integral of this local force over the entire bead surface:

$$\mathbf{F}_a(t) = \int dS \; g(\mathbf{r}, t) \; \hat{\mathbf{n}}(\mathbf{r}). \tag{1.1}$$

2. In response to this force, the bead moves with an instantaneous velocity $\mathbf{v}(t)$. The velocity is linearly related to the force via

$$v_\alpha(t) = \mu_{\alpha\beta} F_\beta(t) = \mu_{\alpha\beta} \int dS\ g(\mathbf{r},t)\ n_\beta(\mathbf{r}), \tag{1.2}$$

where $\alpha, \beta = 1, 2, 3$ and the mobility tensor μ depends on the shape of the object. We are unconcerned with the exact nature of the drag force associated with this mobility[1]. We only assume that at small enough velocities we may make an expansion for velocity in terms of force, with Eq. (1.2) representing the leading term of this expansion. At this stage we neglect rotational motion of the bead.

3. We assume that the rate of change of $g(\mathbf{r},t)$ is an unknown function Φ of the local value of g and the bead velocity $\mathbf{v}$. The dependence on $\mathbf{v}$ couples the field g at different positions on the bead and provides the possibility of positive feedback. We expand the function Φ in powers of g and $\mathbf{v}$, resulting in

$$\begin{aligned}\frac{\partial g(\mathbf{r},t)}{\partial t} = \Phi\left[g(\mathbf{r}),\mathbf{v}\right] = &-g + a\mathbf{v}\cdot\hat{\mathbf{n}} - g^2 + bg\mathbf{v}\cdot\hat{\mathbf{n}} \\ &+d\mathbf{v}\cdot\mathbf{v} + e\left(\mathbf{v}\cdot\hat{\mathbf{n}}\right)^2 - cg^3 + \cdots + \eta(\mathbf{r},t),\end{aligned} \tag{1.3}$$

where we have included all terms to second order and the simplest cubic term. Note that the vector $\mathbf{v}$ appears through the scalar forms $v_\perp = \mathbf{v}\cdot\hat{\mathbf{n}}$ and $v^2 = \mathbf{v}\cdot\mathbf{v}$. This expansion is presumably valid close to a continuous transition between stationary and moving states, during which both g and $\mathbf{v}$ are small; the existence of such a continuous transition needs to be self-consistently justified. We have rescaled g and t so that the coefficients of g and g^2 are both -1. The sign of the linear term is negative so that a uniform distribution of actin (set to $g = 0$ without loss of generality) is stable in the absence of any velocity coupling, while the sign of the g^2 term is unimportant since we may always consider the dynamics of $-g$ rather than g. We are also free to rescale lengths, and

[1]For functionalized beads propelled by an actin comet tail, there is evidence that drag from tethered filaments far exceeds viscous drag due to the surrounding medium [33]. Whether this is true for a stationary bead in an actin cloud is unknown.

a convenient choice is to set the surface area of the bead to unity. Through appropriate redefinitions, our previous formulas for the force $\mathbf{F}_a(t)$ and bead velocity $\mathbf{v}(t)$ remain unchanged. For analysis and numerical simulations of Eq. (1.3) we typically set $d = e = \cdots = 0$, and $c > 0$ for stability. The final term in Eq. (1.3) allows for a stochastic noise $\eta(\mathbf{r}, t)$.

1.3 Analysis

1.3.1 Spherical Beads

We first analyze the case of a spherical bead, and then generalize its lessons to more complicated shapes. Due to its symmetry, the velocity and force are parallel for a sphere, with $\mu_{\alpha\beta} = \mu\delta_{\alpha\beta}$. In the absence of noise η, the dynamics admit the trivial solution of $g(\mathbf{r}, t) = \mathbf{v}(t) = 0$ which corresponds to a static, uniform actin distribution and a motionless bead. This solution is linearly stable for $a < 3/\mu$, but unstable to dipolar fluctuations for $a > 3/\mu$. In the latter case, the unstable actin fluctuations grow and saturate. Thus our model describes a symmetry-breaking transition from a bead at rest ($\mathbf{v} = 0$) to a bead in motion ($\mathbf{v} = const$) as the parameter a is varied past a critical value $a_c = 3/\mu$. This parameter controls the strength of the positive feedback in the coupling of the bead velocity to the actin field.

Near this transition we expand g in spherical harmonics with $g(\mathbf{r}, t) \to g(\Omega, t) = \sum_{\ell m} g_{\ell m}(t) Y_{\ell m}(\Omega)$, where Ω represents the solid angle coordinates θ and ϕ. In this basis Eq. (1.3) becomes

$$\frac{dg_{\ell m}}{dt} = \left(-1 + \frac{\mu a}{3}\delta_{\ell 1}\right) g_{\ell m} + \text{nonlinear terms}, \tag{1.4}$$

where the nonlinear terms couple different $g_{\ell m}$'s with coefficients of the form $\int d\Omega Y_{\ell_1 m_1} Y_{\ell_2 m_2} Y^*_{\ell m}$, $\int d\Omega Y_{\ell_1 m_1} Y_{\ell_2 m_2} Y_{\ell_3 m_3} Y^*_{\ell m}$, etc. From the linear terms in Eq. (1.4) we see that, near transition with $\epsilon \equiv (a - a_c)/a_c \ll 1$ and for small actin fluctuations, the dipolar ($\ell = 1$) modes evolve on a slow time scale $t \propto |\epsilon|^{-1}$, while the other modes evolve on a fast time scale $t \propto \mathcal{O}(1)$. Due to their instability (or near instability) we also

expect the dipolar modes to be larger than the others. We therefore treat the fast $\ell \neq 1$ modes as adiabatically slaved to the slow $\ell = 1$ modes and self-consistently solve for their amplitudes as functions of the g_{1m}'s. Substitution then yields evolution equations for the slow modes, and by a simple transformation, the bead velocity[2]:

$$\frac{d\mathbf{v}}{dt} = \epsilon \mathbf{v} + \frac{27}{5\mu^2}\left[\left(\frac{\mu b}{3} - 1\right)\left(\frac{\mu b}{3} - 2\right) - c\right]\mathbf{v}^3 + u\mathbf{v}^5 + \cdots, \tag{1.5}$$

with $\mathbf{v}^3 \equiv v^2\mathbf{v}$, etc. Near transition the bead velocity $\mathbf{v}$ obeys a Ginzburg-Landau equation consistent with spherical symmetry. When the coefficient of the cubic term is negative, the transition to nonzero $\mathbf{v}$ is continuous with the steady state velocity $\mathbf{v}^*$ going to zero as $\epsilon \to 0^+$. In the vicinity of this continuous transition our expansion Eq. (1.3) is self-consistently justified and Eq. (1.5) provides an accurate description of the dynamics. The fifth-order coefficient u, although not explicitly presented, is negative in this region, providing overall stability.

The phase diagram and critical behavior associated with the steady-state solutions $\mathbf{v}^*$ of Eq. (1.5) are familiar from mean-field theory in statistical mechanics. The cubic coefficient is negative when $b_{c1} < b < b_{c2}$, where b_{c1} and b_{c2} are the two roots of $(\mu b/3-1)(\mu b/3-2) = c$ (recall that $c > 0$). When b lies in this interval, the transition at $\epsilon = 0$ is continuous with $\mathbf{v}^* \propto \epsilon^{1/2}$ for $\epsilon > 0$. When b lies outside this interval, the cubic coefficient is positive and the transition is discontinuous. Tricritical points occur when $b = b_{c1}$ or b_{c2} precisely. At these points the cubic term vanishes and $\mathbf{v}^* \propto \epsilon^{1/4}$. We may also calculate the critical behavior through the tricritical points along the line $\epsilon = 0$ or $a = a_c$. Along this line the linear term in Eq. (1.5) vanishes and $\mathbf{v}^* \propto |b - b_{ci}|^{1/2}$ for $i = 1, 2$ as b varies outside of the interval $b_{c1} < b < b_{c2}$.

Since the transition at $a = a_c$ is discontinuous for $b > b_{c2}$ or $b < b_{c1}$, in these regions the $\mathbf{v} = const$ solutions that exist for $a > a_c$ also exist and are locally stable for $a^* < a < a_c$, where the limit of metastability a^* has some b dependence. Near the tricritical points we find that $a_c - a^* \propto (b - b_c)^2$ for $b < b_{c1}$ or $b > b_{c2}$. In the region between the limit of metastability a^* and the line $a = a_c$, both the $\mathbf{v} = 0$ and the

[2]Equation (1.5) neglects terms of the form $\epsilon^n \mathbf{v}^3$ with $n = 1, 2, 3, \ldots$ and similarly at order $\mathbf{v}^5$. Calculation of these terms by adiabatic elimination requires more care.

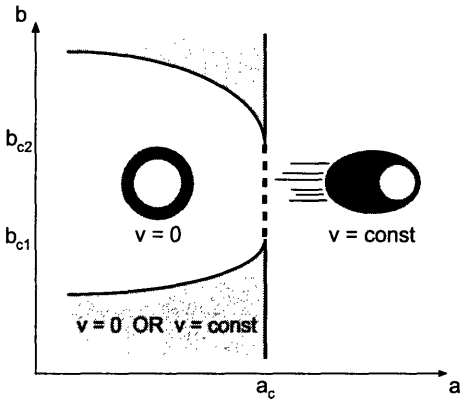

Figure 1-2: Phase diagram in the a–b plane for the motion of a spherical bead in the vicinity of continuous transitions at $a = a_c$ and $b_{c1} < b < b_{c2}$.

(infinitely degenerate) $\mathbf{v} = const$ fixed points are locally stable. If there is a source of noise, the system can in principle switch back and forth between the two states. The phase diagram in Fig. 1-2 summarizes the above results.

To confirm the analytical results, we numerically simulated Eq. (1.3) with a fourth-order Runge-Kutta scheme. For integrations over the surface of a sphere, we adapted routines from the NAG library [2]. We confirmed the phase diagram in Fig. 1-2, as well as the various critical exponents. As an example of the latter, Fig. 1-3 illustrates the time evolution of velocity for $a > a_c = 1$. Starting from an initial (unstable) stationary state, any small fluctuation takes the sphere into a moving phase. The characteristic time scale for this change of state diverges as $\epsilon^{-1} \propto 1/(a - a_c)$, while the saturation velocity scales as $\epsilon^{1/2} \propto \sqrt{a - a_c}$. As indicated in Fig. 1-3, the velocity evolution curves obtained for a number of different a can be collapsed using these scalings of velocity and time.

1.3.2 Nonspherical Beads

We now consider a rigid bead of arbitrary shape. The analog of Eq. (1.5) for the sphere can be obtained by taking a time derivative of Eq. (1.2) and substituting from Eq. (1.3):

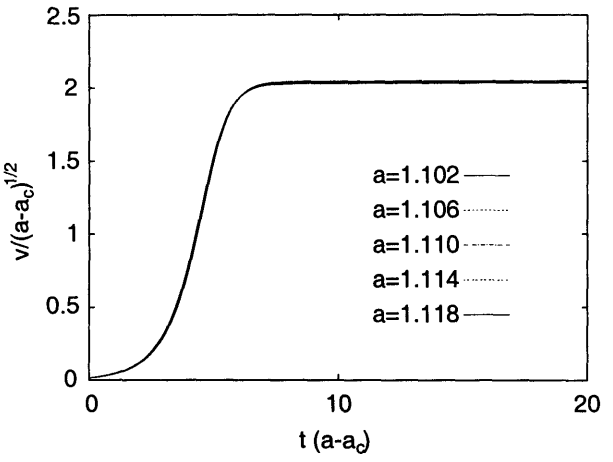

Figure 1-3: Rescaled bead speed versus rescaled time for $\mu = 3$, $c = 0.2$, $b = 1.5$ ($b_{c1} \approx 0.83$ and $b_{c2} \approx 2.17$), and various values of $a > a_c = 1$. The bead begins (very nearly) at rest at $t = 0$ with a small random actin distribution. The curves are nearly perfectly superimposed.

$$\frac{dv_\alpha}{dt} = (-\delta_{\alpha\beta} + a\ \mu_{\alpha\gamma}\ \overline{n_\gamma n_\beta}\ )\, v_\beta + \cdots . \tag{1.6}$$

At linear order, properties of the bead shape are encoded in the tensor $\overline{n_\alpha n_\beta} \equiv \int dS\ n_\alpha(\mathbf{r}) n_\beta(\mathbf{r})$. The linear stability of a stationary bead is determined by the eigenvalues $\{\lambda_i\}$ of the matrix $\mu \cdot \overline{\mathbf{nn}}$. When $a\lambda_i > 1$ for some i, the bead is unstable to motion in the corresponding eigendirection. The largest eigenvalue indicates the direction that initially becomes unstable. Of course, as in Fig. 1-2, it is possible to have coexistence of moving and stationary states before the onset of this instability. In fact, the analysis so far does not rule out coexistence of several states moving along different directions.

It is interesting to note that the eigenvalues of $\overline{\mathbf{nn}}$ are larger along directions where the shape displays a bigger cross-section. Thus, ignoring variations in mobility, a pancake shape prefers to move along its axis (breaking a two-fold symmetry), while a cigar shape moves perpendicular to its axis (breaking a degeneracy in angle). In contrast, for passive ellipsoids in a fluid, hydrodynamic effects favor motion along the slender axis. Experiments on functionalized glass rods [8] and polystyrene disks [29] agree with this prediction of our model. On the other hand, *Listeria* apparently contradicts the theory by moving along its slender axis. This discrepancy is likely

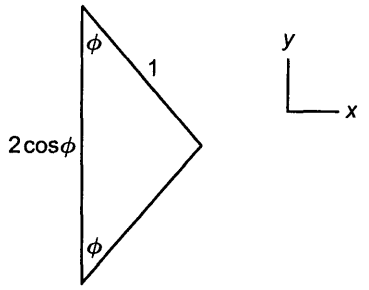

Figure 1-4: An active triangular bead moving in two dimensions. For $\phi < \frac{\pi}{3}$ the bead becomes spontaneously motile in the x–direction first as a is increased.

due to the asymmetric coating of ActA on the surface of *Listeria* [19].

To consider Eq. (1.6) beyond linear order, we adiabatically eliminate all modes of $g(\mathbf{r}, t)$ except the three-dimensional subspace corresponding to the three macroscopic velocities of the bead[3]. If the transition at $a = a_c$ is continuous, its singularities are determined by the next higher-order term in this equation. For the sphere, symmetry forbids a quadratic term; the cubic term leads to the singularities discussed earlier. However, quadratic terms $v_\beta v_\gamma$ may appear in Eq. (1.6) with coefficients $C_{\alpha\beta\gamma}$ that are related to the shape by $\overline{n_\alpha n_\beta n_\gamma} \equiv \int dS\, n_\alpha(\mathbf{r}) n_\beta(\mathbf{r}) n_\gamma(\mathbf{r})$. If such terms are present, the velocity will vanish on approaching the transition as ϵ, rather than $\sqrt{\epsilon}$. To distinguish the two universality classes along a particular direction, we merely need to ask if the opposite direction is equivalent. Thus the velocity of a cigar-shape will vanish as $\sqrt{\epsilon}$, while that of an arrowhead goes to zero linearly. In the language of dynamical systems, the transition at $\epsilon = 0$ is a pitchfork bifurcation in the former case and a transcritical bifurcation in the latter [31].

As an example of the transcritical case, consider the triangular bead depicted in Fig. 1-4. We neglect rotational motion and for convenience eliminate the $-g^2$ term in the actin dynamics Eq. (1.3). (As before we set $c = 1$ and $d = e = \cdots = 0$.) For $\phi < \frac{\pi}{3}$, the bead becomes unstable to motion in the x–direction first for increasing a, with $a_c = (2\cos\phi + 2\cos^2\phi)^{-1}$. Near transition v_x obeys

[3]This is only reasonable near transition when $|a\lambda_1 - 1| \ll 1$, where λ_1 is the largest eigenvalue of $\mu \cdot \overline{\mathbf{nn}}$. The elimination of fast modes is only partial, in that the components of $\mathbf{v}$ in directions perpendicular to the unstable eigenvector are fast, but still retained.

$$\frac{dv_x}{dt} = \epsilon v_x - 2b \sin^2\left(\frac{\phi}{2}\right) v_x^2 + \cdots, \tag{1.7}$$

while $v_y = 0$ is stabilized at the linear level on a fast time scale. The triangular bead does not possess symmetry in the x–direction that prohibits a quadratic term from appearing in the dynamics, and here we have explicitly calculated this term. For $a > a_c$, the triangular bead becomes spontaneously motile to the right or left for $b > 0$ or $b < 0$ respectively, with a steady-state velocity v_x^* that goes linearly with ϵ.

1.3.3 Rotational Motion

We now include the rotational motion of the bead in our analysis. We assume a linear dependence of the bead's center of mass velocity $\mathbf{V}$ and angular velocity $\boldsymbol{\omega}$ on the external force $\mathbf{F}$ and torque $\mathbf{N}$ experienced by the bead:

$$\begin{pmatrix} \mathbf{V} \\ \boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \mu^{VF} & \mu^{VN} \\ \mu^{\omega F} & \mu^{\omega N} \end{pmatrix} \begin{pmatrix} \mathbf{F} \\ \mathbf{N} \end{pmatrix} = \mathbb{M} \begin{pmatrix} \mathbf{F} \\ \mathbf{N} \end{pmatrix}, \tag{1.8}$$

where $\mathbb{M}$ is a 6×6 mobility matrix[4]. As before, the activity at a point $\mathbf{r}$ on the bead surface generates a local force density $g(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})$, the net force being the integral of this quantity over the entire bead surface (recall Fig. 1-1). In addition we now consider the local *torque* density $g(\mathbf{r})\mathbf{r} \times \hat{\mathbf{n}}(\mathbf{r})$ and corresponding net torque $\mathbf{N}$. To complete the generalization of our model, we replace the bead velocity $\mathbf{v}$ that appears in the microscopic dynamics Eq. (1.3) with the *local* velocity of the bead surface at $\mathbf{r}$, i.e. we put $\mathbf{v} \to \mathbf{v}(\mathbf{r}) = \mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}$. Eq. (1.3) becomes

$$\frac{\partial g(\mathbf{r})}{\partial t} = -g(\mathbf{r}) + a(\mathbf{V} + \boldsymbol{\omega} \times \mathbf{r}) \cdot \hat{\mathbf{n}}(\mathbf{r}) + \text{nonlinear terms}. \tag{1.9}$$

Taking the time derivative of Eq. (1.8) in a bead-fixed frame, we find

[4]For a rigid object in a fluid at low Reynolds number, the forms of the various mobilities μ are constrained. See, for example, [15].

$$\left(\frac{\partial}{\partial t}\right)_{bead} \begin{pmatrix} \mathbf{V} \\ \boldsymbol{\omega} \end{pmatrix} = \mathbb{Q} \begin{pmatrix} \mathbf{V} \\ \boldsymbol{\omega} \end{pmatrix} + \text{nonlinear terms}, \tag{1.10}$$

where

$$\mathbb{Q}_{ij} = -\delta_{ij} + a\mathbb{M}_{ik}\overline{B_k B_j}; \qquad B = \begin{pmatrix} \hat{\mathbf{n}} \\ \mathbf{r} \times \hat{\mathbf{n}} \end{pmatrix}. \tag{1.11}$$

Here the Latin indices run from 1 through 6, and as before the overline indicates an average over the bead surface, i.e. $\overline{f(\mathbf{r})} = \int dS f(\mathbf{r})$. The form of Eq. (1.11) is the same as that of Eq. (1.6). In the stationary state $\mathbf{V} = 0$, $\boldsymbol{\omega} = 0$, all modes are stabilized by the Kronecker delta, while some are destabilized by actin-based feedback, the strength of which is governed by the parameter a. As a is increased, eigenmodes of $\mathbb{Q}$ progressively become unstable, with the largest eigenvalue determining the limit of stability of the stationary state. Near transition, the nonlinear terms in Eq. (1.10) may be computed by eliminating fast modes. As before, we expect the form of these nonlinear terms to respect symmetry constraints.

For objects with relatively high symmetry, the tensor $\mathbb{M}\overline{BB}$ often takes a diagonal form. For a thin rod oriented along the 1-direction, for example, $(\mathbb{M}\overline{BB})_{ij}$ is nonvanishing only for $i = j = 2, 3, 5$, or 6, corresponding to translational motion in the 2–3 plane or rotational motion through axes in the 2–3 plane. For a planar object such as a thin ellipsoidal disk with the normal pointing along the 1-direction, $(\mathbb{M}\overline{BB})_{ij}$ is nonvanishing only for $i = j = 1, 5$, or 6, corresponding to translational motion in the 1-direction or rotational motion through axes in the 2–3 plane. For a fully three-dimensional object such as an ellipsoid, $\mathbb{M}\overline{BB}$ is diagonal with nonvanishing entries for all six modes. Note that as before, there is no reason a priori to rule out the coexistence of several different modes of motion.

As a simple illustration, we consider a thin rod with length $\ell = 2$ fixed at its center and free to rotate in a plane. See Fig. 1-5. We put $d = e = \cdots = 0$ in the microscopic dynamics Eq. (1.3) and take the rotational mobility ν to be unity. For $a > a_c = 3/4$ the rod is unstable to spontaneous rotations; the unstable mode of g is

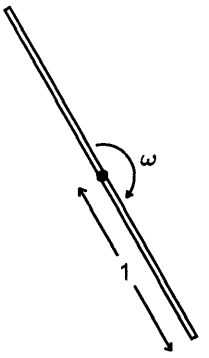

Figure 1-5: Thin rod rotating in two dimensions about a fixed point at its center. The actin field $g(\mathbf{r}, t)$ is defined along the two long edges of the rod.

a linear function of the distance from the rod's center. Near transition, ω obeys

$$\frac{d\omega}{dt} = \epsilon\omega + \frac{3}{5}\left[\left(b - \frac{3}{4}\right)\left(b - \frac{3}{2}\right) - \frac{9}{16}c\right]\omega^3 + \cdots, \tag{1.12}$$

where $\epsilon \equiv (a - a_c)/a_c$. This equation is similar in form to Eq. (1.5) for the spherical bead and should result in a similar phase diagram in the a–b plane.

We may generalize the rod problem slightly by allowing the rod to translate as well as rotate. Since the rod is thin, there is no activity at its ends and the rod only translates in a direction perpendicular to its axis, say with center of mass velocity V. This is the simplest situation involving both translations and rotations, with a single degree of freedom for each. Generically, the rod becomes unstable to one type of motion before the other as a is increased. (As usual, there may be coexistence in some regions of the phase diagram.) As a theoretical exercise we may ask what happens if we tune the system so that the two modes become unstable simultaneously. This can be achieved, for example, by manipulating the linear and rotational mobilities. As in the case of the triangular bead we use the microscopic dynamics Eq. (1.3) with no $-g^2$ term, $c = 1$, and $d = e = \cdots = 0$. We find near transition coupled dynamics of the form

$$\frac{d\omega}{dt} = \epsilon\omega + \frac{3}{5}s\omega^3 + 3s\omega V^2$$

$$\frac{dV}{dt} = \epsilon V + sV^3 + sV\omega^2, \tag{1.13}$$

where $a_c \equiv (4\mu)^{-1}$, $\epsilon \equiv (a - a_c)/a_c$, $s = s(\mu, b) = b^2 - (16\mu^2)^{-1}$, and the mobilities ($\mu$ for translational motion and ν for rotational motion) are tuned so that $\nu = 3\mu$, which gives the simultaneous criticality. A second-order transition at $a = a_c$ requires $s < 0$, which occurs when $-(4\mu)^{-1} < b < (4\mu)^{-1}$. Rescaling ω and V appropriately yields dynamics that are potential, with an energy function of the form

$$F(\omega, V) = -\frac{1}{2}\epsilon(\omega^2 + V^2) + \frac{1}{4}A\omega^4 + \frac{1}{4}BV^4 \pm \frac{1}{2}\omega^2 V^2. \tag{1.14}$$

In fact, from symmetry constraints on the macroscopic dynamics we see that the system is always governed by a potential function of this form. Just above a second-order transition ($0 < \epsilon \ll 1$), Eq. (1.14) has possible fixed points at $|\omega| \neq 0$ and $|V| = 0$, $|\omega| = 0$ and $|V| \neq 0$, and $|\omega| \neq 0$ and $|V| \neq 0$. In general, the stability of these fixed points depends on the coefficients A and B, and on the sign of the coupling term. The somewhat surprising point is that, regardless of the values of the parameters $b, c, d, \cdots$ in the actin dynamics Eq. (1.3), we always find that $0 < A, B < 1$ and that the coupling term has positive sign. The result is that the $|\omega| \neq 0$, $|V| \neq 0$ fixed point is always unstable, while the other fixed points are always stable. A thin rod moving in two dimensions cannot stably rotate and translate simultaneously near transition.

This observation leads to the interesting theoretical issue of determining what sorts of objects, if any, permit such a stable combination of translational and rotational motion. A propeller-shaped object seems like a likely candidate. Such an object, however, is "naturally" inclined to simultaneously rotate and translate, i.e. the off-diagonal mobilities μ^{VN} and $\mu^{\omega F}$ in Eq. (1.8) are nonvanishing (for any coordinate system fixed in the bead). It would be interesting to determine whether an object without this natural inclination, i.e. with $\mu^{VN} = \mu^{\omega F} = 0$, could be driven to simultaneously rotate and translate at transition purely from actin dynamics. From a purely mathematical standpoint, there is no reason to believe that this situation is

impossible. Mathematically the situation in which V and ω become simultaneously critical is very similar to the situation, in the case of a sphere in two dimensions for example, in which v_x and v_y become simultaneously critical. In the case of a sphere, $A = B = 1$ in the potential analogous to Eq. (1.14); the two modes v_x and v_y can freely mix at transition, leading to symmetry-breaking in an arbitrary direction. The issue in the case of simultaneous translation and rotation is whether there are general geometric constraints that limit, for example, the possible values of A and B to a certain regime.

1.3.4 Beads in External Potentials

We now consider an active bead moving in a potential well. For simplicity, we constrain the bead to move in the x-direction and ignore rotational motion. The net force on the particle (excluding drag) is

$$F(t) = F_a(t) + F_{pot}(t) = \int dS\ g(\mathbf{r}, t) n_x(\mathbf{r}) - \phi'(x). \tag{1.15}$$

The bead velocity is given by $v(t) = \mu F(t)$. Eq. (1.15) leads to the linearized dynamics

$$\begin{aligned} \dot{x} &= v \\ \dot{v} &= -[r + \phi''(x)]v - \phi'(x), \end{aligned} \tag{1.16}$$

where $r \equiv 1 - a\mu\overline{n_x n_x}$ and $\phi(x)$ has been redefined to absorb a factor of μ. At the linear level only the harmonic component of the potential $\phi(x)$ contributes through the terms $-\phi''(x)v$ and $-\phi'(x)$. Eqs. (1.16) are the same as those describing a particle moving through a viscous medium in a harmonic potential, except for the fact that the "drag" $r + \phi''$ can become negative for large enough a. The stationary state becomes unstable when $a > [1 + \phi''(0)]/\mu\overline{n_x n_x}$. The bifurcation at $a = a_c$ is Hopf so that for $a > a_c$ the system is unstable to an oscillating mode. This is consistent with the intuition of a particle in a harmonic well with "negative drag". Of course,

nonlinearities in the dynamics that are neglected in Eq. (1.16) can control these growing oscillations.

1.4 Fluctuations

Our objective in this section is to make some observations on the effect of the noise term η in the actin dynamics Eq. (1.3). This term is intended to account for all manner of stochasticity in the polymerization and force generation process, including, for example, fluctuations in the concentration of monomeric actin or other relevant proteins, the intrinsic stochasticity of the polymerization process, disorder in the cross-linked actin network, etc. In the absence of a microscopic model, a detailed account of the origin or form of η cannot be given. For simplicity in the analysis we take η to be delta-correlated in space and time with $\langle \eta(\mathbf{r},t)\rangle = 0$ and $\langle \eta(\mathbf{r},t)\eta(\mathbf{r}',t')\rangle = 2A\delta^2(\mathbf{r}-\mathbf{r}')\delta(t-t')$, the parameter A determining the noise strength.

1.4.1 Spherical Beads

We begin by investigating the effect of noise on spherical beads in the $\mathbf{v} = 0$ phase. We approximate the dynamics in Eq. (1.3) by its linear terms, yielding

$$\frac{\partial g(\mathbf{r},t)}{\partial t} = -g(\mathbf{r},t) + a\mathbf{v}(t)\cdot\hat{\mathbf{n}}(\mathbf{r}) + \eta(\mathbf{r},t). \tag{1.17}$$

The deterministic part of the dynamics is potential, with an energy function

$$\mathcal{H}[g(\mathbf{r})] = \frac{1}{2}\int dS g^2 - \frac{1}{2}a\mu \int dS_1 dS_2 g(\mathbf{r}_1)g(\mathbf{r}_2)\hat{\mathbf{n}}(\mathbf{r}_1)\cdot\hat{\mathbf{n}}(\mathbf{r}_2). \tag{1.18}$$

The steady-state probability distribution for g is given by $P^*[g] \propto e^{-\beta\mathcal{H}[g]}$, with $\beta = A^{-1}$; the corresponding distribution for $\mathbf{v}$ is given by $P^*(\mathbf{v}) \propto e^{-\gamma v^2}$ with $\gamma = \frac{1}{2A\mu}(a_c - a)$. Actin correlations at long times are given by

$$\langle g(\Omega,t)g(\Omega',t')\rangle_c \propto A\delta^2(\Omega-\Omega')e^{-|t-t'|} + \frac{3}{4\pi}A\cos\gamma\left(\tau e^{-|t-t'|/\tau} - e^{-|t-t'|}\right), \tag{1.19}$$

where $\tau = |\epsilon|^{-1}$ is the relaxation time for dipolar actin fluctuations and γ is the angle between Ω and Ω'. Actin fluctuations on the same side (opposite sides) of the bead are positively (negatively) correlated. Velocity correlations for the bead are given by $\langle \mathbf{v}(t) \cdot \mathbf{v}(t') \rangle \propto A\tau e^{-|t-t'|/\tau}$.

Analysis of fluctuations in the $\mathbf{v} = const$ phase appears more difficult. Qualitatively we expect that for weak levels of noise the bead will move with a nearly constant speed, while its direction of motion evolves like a random walk. At long times and distance scales the motion is diffusive, while at short times $t \lesssim A^{-1}$ the motion is ballistic. In the coexistence region, noise can drive the system between the $\mathbf{v} = 0$ and the degenerate manifold of $\mathbf{v} = const$ states.

1.4.2 Non-Maxwellian Fluctuations

For the remainder of the analysis on fluctuations, we focus on the issue of how the steady-state velocity distributions for an active bead in the stationary $\mathbf{v} = 0$ phase differ from those for a passive Brownian particle. If rotational motion is forbidden, we may perform a simple analysis by including the noise in Eq. (1.6) and keeping only the linear terms. The resulting Langevin equation is given by

$$\dot{v_\alpha} = K_{\alpha\beta} v_\beta + f_\alpha, \tag{1.20}$$

where $K_{\alpha\beta} = -\delta_{\alpha\beta} + a\mu_{\alpha\gamma}\overline{n_\gamma n_\beta}$, and the noise f is characterized by

$$\langle f_\alpha(t) \rangle = 0; \qquad \langle f_\alpha(t) f_\beta(t') \rangle = 2A\delta(t-t')\mu_{\alpha\gamma}\mu_{\beta\kappa}\overline{n_\gamma n_\kappa} \equiv Q_{\alpha\beta}\delta(t-t'). \tag{1.21}$$

The steady-state velocity distribution in this case is Gaussian, $P^*(\mathbf{v}) \propto \exp\left(v_\alpha M^{-1}_{\alpha\beta} v_\beta/2\right)$, where the variances $M_{\alpha\beta}$ can be found by solving the matrix equation $KM + MK^T = -Q$ [27]. In contrast to a passive Brownian particle, whose velocity distribution is Maxwellian, the matrix $M_{\alpha\beta}$ is in general not proportional to $\delta_{\alpha\beta}$, i.e. the fluctuations of an active particle are correlated with the particle shape and may be larger in some

directions than others. For an object where $\mu_{\alpha\beta}$ and $\overline{n_\alpha n_\beta}$ are diagonal, for example, we have $M_\alpha = A\mu_{\alpha\alpha}^2\overline{n_\alpha n_\alpha}/(1-a\mu_{\alpha\alpha}\overline{n_\alpha n_\alpha})$ (no summation). Whether the fluctuations of, for example, a pancake-shaped object are larger along its normal axis or along an axis lying in its plane depend on the precise forms of μ and $\overline{nn}$. The difference between the active and passive cases is that for a Brownian particle, fluctuation–dissipation constraints enforce a particular relationship between the viscous drag and the noise that leads to the isotropic velocity distribution. For the active particle this relationship is spoiled.

The same situation occurs for rotational motion. Consider an arbitrarily shaped active bead free to rotate in three dimensions about its center of mass. The linearized rotational dynamics (see Eqn. (1.10)) leads to the Langevin equation

$$\dot{\omega}_i = K_{ij}\omega_j + f_i, \tag{1.22}$$

where now $K_{ij} = -\delta_{ij} + a\nu_{ik}\overline{(\mathbf{r}\times\hat{\mathbf{n}})_k(\mathbf{r}\times\hat{\mathbf{n}})_j}$, with ν the rotational mobility tensor. The noise f is characterized by

$$\langle f_i(t)\rangle = 0; \qquad \langle f_i(t)f_j(t')\rangle = 2A\delta(t-t')\nu_{ik}\nu_{jl}\overline{(\mathbf{r}\times\hat{\mathbf{n}})_k(\mathbf{r}\times\hat{\mathbf{n}})_l} \equiv Q_{ij}\delta(t-t'). \tag{1.23}$$

The similarity with the translational case is evident. For, say, an ellipsoidal particle, the tensors K and Q are diagonal in the principal coordinate system, i.e. the dynamics are decoupled into rotations along the principal axes. The steady-state solution to the Fokker-Planck equation for this system [18] is Gaussian with $P^*(\boldsymbol{\omega}) \propto \exp(-M_i\omega_i^2/2)$. The inverse variances M_i are not in general proportional to the moments of inertia I_i of the particle, as would be the case for a rotating Brownian particle in thermal equilibrium. Indeed, for an active particle the moments of inertia are in fact irrelevant to the dynamics since we are considering the regime of overdamped motion.

As a final example, we reconsider an active particle moving in an external potential. As in Sec. 1.3.4, we consider a bead moving in one dimension without rotating.

The dynamics are given by Eq. (1.16) with the addition of a noise term f in the velocity dynamics. We are interested in the steady-state distribution $P^*(x,v)$ for the position and velocity of the bead. Using the Fokker-Planck equation for this system and setting $\partial P/\partial t = 0$ leads to

$$-v\partial_x P + (r + \partial_x^2\phi)(v\partial_v P + P) + (\partial_x\phi)(\partial_v P) + Q\partial_v^2 P = 0, \tag{1.24}$$

where Q is given by the variance of the noise f, $Q \equiv A\mu^2\overline{n_x n_x}$. For the case of a Brownian particle with unit mass, the equilibrium distribution is given by $P^*(x,v) \propto \exp[-\beta(v^2/2 + \phi(x))] = P^*(x)P^*(v)$. In this case the particle position and velocity are independent and covariances like $\langle xv\rangle_c$ trivially vanish. Eq. (1.24), on the other hand, does not in general admit a factorable solution of the form $P^*(x)P^*(v)$, doing so apparently only for harmonic potentials. For anharmonic potentials one expects nonvanishing position-velocity correlations; in particular, one expects that $\langle xv\rangle_c = 0$ only by accident.

To verify this expectation, we solved Eq. (1.24) numerically using finite differencing [26] for various anharmonic potentials. In all cases tried, we found that $\langle xv\rangle_c < 0$. Fig. 1-6 gives a typical result of this calculation.

1.5 Conclusions and Discussion

Motivated by experiments on functionalized spherical beads and their actin-based symmetry-breaking motility, we have developed a phenomenological model for the active motility of rigid particles. In our model a motile object is propelled by forces at its surface generated by some energy-consuming process such as actin polymerization. The dynamical evolution of these forces depends locally on the velocity of the bead surface through the surrounding medium. This velocity coupling can provide positive feedback so that the bead can become unstable to one or more modes of motion. Spontaneous motility and symmetry-breaking are interpreted as consequences of this instability. The main idea of our approach is to make an expansion for the force dynamics near the instability threshold and to eliminate fast microscopic modes,

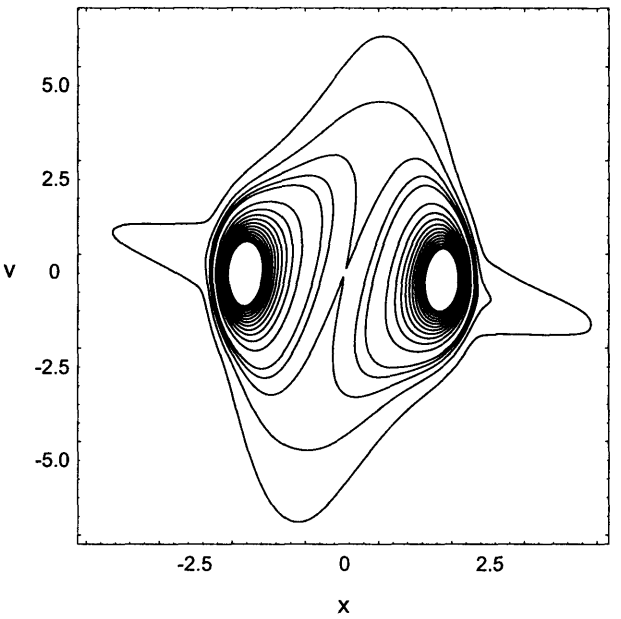

Figure 1-6: Numerically calculated solution $P^*(x, v)$ to Eqn. (1.24) for the steady-state probability distribution of an active bead in a one-dimensional potential well. Here $r = 1$, $Q = 5$, and the potential has a double-well shape with $\phi(x) = \frac{1}{4}x^4 - x^2$. The contours are lines of equal probability. The distribution is peaked at the two minima of ϕ. Note that $P^*(x, v)$ does not possess any symmetry that would cause $\langle xv \rangle$ to vanish.

leading to equations of motion for the macroscopic observables, the translational and rotational velocities of the bead.

The main conclusions of our study are as follows. Near transition, we find that the macroscopic velocities of the object obey Ginzburg-Landau equations consistent with the symmetries of the object. These equations determine the phase diagram for the macroscopic behavior of the bead near transition and provide the various singularities encountered. These singularities (or more generally, the bifurcation structure) are dependent on symmetries in the object's shape. We find that the fluctuations in velocity for an active particle in the stationary phase differ qualitatively from the Brownian fluctuations of a passive particle with the same shape.

Our model has its weaknesses when viewed from the standpoint of describing the experimentally observed phenomena discussed in the introduction. One obvious drawback of our phenomenological approach is that no connection is made to microscopic details. As a result the model has very little to say, for example, on the effects of varying the different experimentally tunable parameters, such as the concentra-

tions of various factors in the medium or the density of the coating on the beads. The model also neglects a major physical aspect of the experimental situation, that of nonlocal interactions within the actin gel. A number of papers [25, 30, 13] have explained symmetry-breaking for spherical beads in terms of a model where the actin cloud is treated as an elastic material that ruptures under stress. The interaction between actin filaments in this situation would be a complicated nonlocal effect that depends on the topology of the bead and gel. On the other hand, there is some experimental evidence for the existence of two regimes, one in which symmetry-breaking is primarily a result of this rupturing effect, and another in which local stochastic fluctuations in polymerization are more important [7]. If this were true, then our model potentially captures the dominant effects for the latter case.

Our theory has other potential weaknesses as well. Strictly speaking, our model applied to actin-based motility describes a bead in an infinite homogeneous and isotropic medium (an actin cloud) capable of exerting force on its surface. This medium responds instantaneously to movement of the bead, making way for the bead in front of it and reforming behind it. The medium carries no memory of its previous history, even if for example the bead has recently traversed through the region in question. In the experimental situation, on the other hand, the actin cloud is plainly finite. When the bead moves towards the edge of the cloud, one expects something to change compared to when the bead is in the middle; in other words one expects the relative position of the bead within the cloud to make some difference. The actin cloud is not isotropic, since polymerization is directional and activated by the bead coating. One would presume that the actin retains some memory of its history as well. The importance of these additional factors in the experimental system that are neglected in the theory is unclear.

We turn our critical eye to the theoretical aspects of our model. As we have noted, the form of the equations of motion for the macroscopic velocities of the bead (e.g. Eqns. (1.5), (1.6), (1.7), etc.) are determined by symmetry. Thus we could have written these equations down immediately without any work at all. One may wonder what exactly is gained by starting from the microscopic Eq. (1.3). The obvious poten-

tial benefit is in being able to calculate the coefficients of the macroscopic equations of motion in terms of experimental parameters; this would require a microscopic model connecting our phenomenological parameters $a, b, c, \cdots$ to experimental factors. In the absence of a microscopic model, the most that can be gained from Eq. (1.3) are constraints for the coefficients in the macroscopic equations. In Sec. 1.3.3 for example, such constraints led to the conclusion that a thin rod moving in two dimensions cannot stably rotate and translate simultaneously near transition. As we speculated in that section, perhaps more general statements of this type can be made. In the end however, most of the mileage from Eq. (1.3) comes from its linear terms $-g + a\mathbf{v} \cdot \hat{\mathbf{n}}$. Diagonalizing the linear dynamics gives the potentially unstable modes and the degree of their instability (i.e. their critical values a_c). The results on non-Maxwellian fluctuations also come from the linear theory alone. Because of the simplicity and generality of the linear dynamics it would not be surprising if aspects of our analysis appear again in other contexts.

Our theory is similar to those used to model pattern formation [9, 3] or "self-organization"[14] in various contexts. These theories typically describe systems in which a uniform, homogeneous state develops a linear instability as a control parameter is varied past a critical value. Above this critical threshold, the instability grows and saturates because of nonlinearities, forming the pattern of interest. Generically, pattern-forming systems are described near threshold by "amplitude equations" that govern the dynamics of the slow, unstable modes and whose form is constrained by symmetries. As in our case, these amplitude equations can often be derived from the microscopic dynamics by eliminating fast modes.

We may view our theory as describing a novel pattern-forming system. In our case the substrate for the pattern is motile and its movement is coupled to the pattern-forming instability. The novelty from a theoretical standpoint is in this coupling, which is geometric in nature. It would be interesting to find other, perhaps biologically-motivated, systems in which geometry and dynamics are so intimately related.

Chapter 2

RNA Secondary Structures: Designability and Combinatorics

2.1 Introduction

RNA is a biopolymer whose most important function in (most) living organisms is to translate the genetic information encoded in DNA into proteins. Like other biopolymers, RNA has structure and organization at several different levels. The *sequence* of an RNA is a string composed of the four letters A, U, G, and C representing the four different nucleotide bases available for incorporation in the backbone. A *base pair* is formed when two bases form hydrogen bonds; chains of consecutive base pairs "stack" to form helices. This level of organization is referred to as the RNA *secondary structure*. Tertiary structure involves the spatial arrangement of the helices of the secondary structure. Since helix-helix interactions are typically significantly weaker than those of base pairing, RNA folding is viewed as hierarchical, and secondary structure is studied in isolation without concern for tertiary interactions.

We will adopt the mathematical definition of RNA secondary structure widely used in the literature. A secondary structure is an undirected graph on N vertices labeled 1 through N representing the bases of the RNA. The edges of the graph represent base pairing; we will refer to these edges either as "base pairs" or "bonds" and denote them by ordered pairs (i, j). Each base can pair at most once, so each

Figure 2-1: RNA secondary structure.

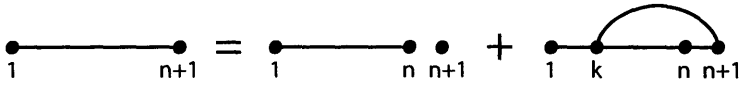

Figure 2-2: Basic recursion to enumerate all RNA secondary structures. The new, $(n+1)$th base is either unpaired or paired.

vertex can have at most one edge. If a secondary structure contains the base pairs (i, j) and (k, l), with $i < j$ and $k < l$ and $i < k < j$, then necessarily $i < k < l < j$. This requirement disallows graphs that contain *pseudoknots*. See Fig. 2-1 for an example of a secondary structure. Forbidding pseudoknots is equivalent to requiring that the arches in this diagram do not cross. Our final restriction is that any base pair (i, j) in a secondary structure must have $|j - i| > d$, where d is the *minimum hairpin gap*. This requirement accounts for the fact that in physical RNAs, steric constraints prevent bases from pairing whenever they are very close to each other along the backbone. The secondary structure in Fig. 2-1, for example, is legal for $d = 2$ but illegal for $d = 3$.

As is well known, it is a simple matter to recursively enumerate all secondary structures of length n. See Fig. 2-2. The same basic recursion can be used to calculate X_n, the number of secondary structures of length n:

$$X_{n+1} = X_n + \sum_{k=1}^{n-d} X_{k-1} X_{n-k}, \quad X_1 = X_2 = \cdots = X_{d+1} = 1 \ . \tag{2.1}$$

Because of this basic recursion, RNA secondary structures are unusually theoretically tractable. For example, algorithms based on dynamic programming techniques have been developed that calculate the exact partition function for an RNA with a given sequence in polynomial time [17]. These algorithms use realistic energy models and rely on experimentally measured physical parameters and consequently have significant predictive power [16].

In this chapter we explore two theoretical topics pertaining to RNA secondary

structures. The first is the sequence to structure map of RNA folding, and in particular, the designability of RNA secondary structures. The second is the statistical mechanics of homopolymer RNAs. Both investigations are incomplete and represent works in progress.

2.2 RNA Folding and Designability

2.2.1 Introduction

A general problem in biopolymer folding is to understand the mapping of sequence to structure. For RNAs, structure prediction for a given sequence has been largely solved by dynamic programming algorithms, so in some sense this map is understood. For proteins, structure prediction remains a difficult problem. It has been suggested that the RNA sequence to structure map is fundamentally different from that of proteins.

We were motivated to explore this issue by an elegant geometric formulation of protein folding developed by Li et al [21]. In this model, the sequence of an N-residue protein is represented by a vector of length N, $\mathbf{S} = (s_1, s_2, \cdots, s_N)$, where s_i is the degree of hydrophobicity of the ith residue. A protein structure is represented by another vector of length N, $\mathbf{F} = (f_1, f_2, \cdots, f_N)$, where $f_i = 0$ for surface sites and $f_i = 1$ for interior or core sites. The allowable structures for a protein are taken to be compact self-avoiding walks on a 2D square lattice. The vectors $\mathbf{F}$ and $\mathbf{S}$ may both be regarded as points in an N-dimensional space.

With these definitions, the energy of a protein with some sequence $\mathbf{S}$ folded into some structure $\mathbf{F}$ can be taken to be $E = -\mathbf{F}\cdot\mathbf{S} = \frac{1}{2}\left[(\mathbf{F}-\mathbf{S})^2 - \mathbf{F}^2 - \mathbf{S}^2\right]$. The *native structure* for a particular sequence $\mathbf{S}^*$ is that structure $\mathbf{F}^*$ for which E is minimized; $\mathbf{S}^*$ is said to *design* $\mathbf{F}^*$. Noting that $\mathbf{F}^2$ is constant over all structures[1], minimization of E for given $\mathbf{S}$ is equivalent to minimization of the squared distance $(\mathbf{F}-\mathbf{S})^2$. Thus the native structure for a given sequence $\mathbf{S}$ is that structure which lies closest (according to the standard Euclidean metric) to $\mathbf{S}$ in the N-dimensional space of

[1] In [21] they considered only structures that fit perfectly into a 6×6 square lattice. In this case, the number of interior sites for any structure is indeed a constant.

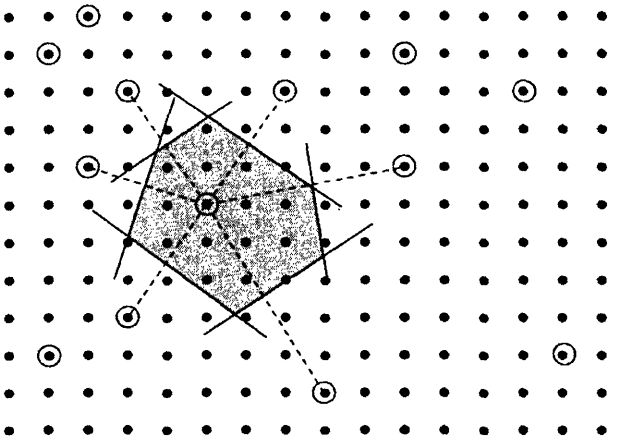

Figure 2-3: Schematic representation of the N-dimensional sequence/structure space. Sequence points are shown as dots, while structure points are shown as circles. One of the structure points is enclosed by its Voronoi polytope. Those sequence points enclosed by the polytope design the structure. Taken from [34].

sequence and structure points. This observation leads to a simple geometric method for finding all sequences that design a particular structure $\mathbf{F}^*$: the volume around $\mathbf{F}^*$ enclosed by drawing bisector hyperplanes between $\mathbf{F}^*$ and all other neighboring structures in sequence/structure space contains all the desired sequence points. This enclosed volume is called the Voronoi polytope. See Fig. 2-3.

One way to gain insight into the mapping of sequence to structure in biopolymer folding is through the notion of *designability* [20]. The designability of a structure is defined to be the number of sequences that design that structure. For the model of Li et al., the fact that the sequence points are uniformly distributed in the sequence/structure space lead via the Voronoi construction to the conclusion that the designability of a structure is proportional to the volume of its Voronoi polytope. Thus the most designable structures are located in regions of low structure point density, while the least designable structures are located in regions of high structure point density. Insight into designability can therefore be gained by investigating the distribution of structure points in the N-dimensional space. Yahyanejad et al. found that the distribution of structure points within this model of protein folding can be well approximated by a Gaussian distribution with a single peak [34].

2.2.2 Model

Our objective is to analyze RNA folding along lines similar to those above. We represent the three-dimensional structure of an RNA N-mer by its adjacency matrix or *contact map*, a symmetric $N \times N$ matrix F in which $F_{ij} = 1$ if the ith and jth residues are in contact and 0 otherwise[2]. In analogy with the contact map, we represent the sequence of an RNA N-mer by an $N \times N$ matrix S in which S_{ij} is the energetic incentive of forming a contact between bases i and j. As an idealized model, we suppose RNA monomers come in two varieties, e.g. purines and pyrimadines. Then, for some RNA sequence $\hat{S} = \hat{s}_1\hat{s}_2\hat{s}_3\cdots\hat{s}_N$ where $\hat{s}_i = 0$ if the ith base is a purine and $\hat{s}_i = 1$ if it is a pyrimidine, we construct the corresponding sequence matrix S by putting $S_{ij} = 1$ if $\hat{s}_i \neq \hat{s}_j$, i.e. if bases i and j are complementary, and 0 otherwise. Note the distinction between an RNA *sequence* $\hat{S}$, which has length N, and an RNA *sequence matrix* S, which is $N \times N$ and exists in an $M = \frac{N(N-1)}{2}$-dimensional space. We will use the term "sequence point" interchangeably with "sequence matrix", and we will call the M-dimensional space the "contact space."

In analogy with Li et al. [21], we assign an energy for folding a sequence $\hat{S}$ into a structure F via $E = -F\cdot S = -F_{ij}S_{ij}$, where S is the sequence point corresponding to the sequence $\hat{S}$. The sequence and structure points S and F may be regarded as points in the same M-dimensional contact space. Reasoning as before, the designability of a particular RNA structure is determined by the local density of sequence and structure points in its neighborhood[3]. Note that in our case, the sequence points are not uniformly distributed in contact space. To the extent that the distributions of the sequence and structure points can be characterized, insight will be gained into the designability of RNA secondary structures.

---

[2]We do not bother to specify a convention for the diagonal entries of F because these entries will not be relevant.

[3]This time, there is a minor technical point in the argument because F^2 is not constant over all structures. We can get around this problem as follows. Instead of the definition we presented, we could have defined $F'_{ij} = 1$ for bases i, j in contact and $F'_{ij} = -1$ otherwise. With *this* definition, F'^2 would indeed be constant. Since $F' = 2F - I$, where $I_{ij} = 1$ for all i, j, $E' = -F'\cdot S$ is minimized when $E = -F\cdot S$ is minimized, for a given sequence point S. Our problem can therefore be avoided by a minor redefinition of F. Since this redefinition doesn't make any important difference as far as an analysis of the distribution of structure points is concerned, we stick to our original definition.

2.2.3 The Distribution of Sequence Points

We begin by investigating the distribution of sequence points S in the M-dimensional contact space for RNAs of length N. For convenience, rather than investigating the distribution of the points S themselves, we investigate the distribution of the inverse points S^* obtained by inverting the points S through the point $(1/2, 1/2, 1/2, \cdots, 1/2)$ in contact space. In other words, we investigate the distribution of the points S^* constructed from RNA sequences $\hat{S}$ by putting $S^*_{ij} = 1$ if $\hat{s}_i = \hat{s}_j$ and $S^*_{ij} = 0$ otherwise[4]. We reiterate that the distributions of the points S^* and the points S in contact space are identical up to inversion. From now on we drop the asterisk and refer only to the sequence points S.

Before proceeding further we make some basic observations. For any RNA sequence $\hat{S}$ and its corresponding sequence matrix S, the inverse sequence $\hat{S}^*$ defined by $\hat{s}^*_i = 0$ if $\hat{s}_i = 1$ and vice versa has an identical sequence matrix. Thus there are a total of 2^{N-1} distinct sequence points in contact space, each one twofold degenerate. There are $2^M = 2^{\frac{N(N-1)}{2}}$ vertices of the unit hypercube in contact space; the sequence points are sparse when considered as a subset of these vertices.

Correlations in the Distribution

One way to investigate how the sequence points are spread throughout contact space is to compute the cumulants $\langle S_{ij} \rangle_c$, $\langle S_{ij} S_{kl} \rangle_c$, $\langle S_{ij} S_{kl} S_{mn} \rangle_c$, etc., where the averages are taken over all sequence points S. The cumulants give the correlations between the components of the sequence points in different coordinate directions in contact space. For the means and covariances we have

$$\langle S_{ij} \rangle_c = 1/2, \quad (i \neq j) \tag{2.2}$$

[4]With this definition the sequence points S^* can be used to model proteins within a simple hydrophobic/polar model. There is positive incentive for hydrophobic/hydrophobic or polar/polar base pairing and no incentive for mixed pairing.

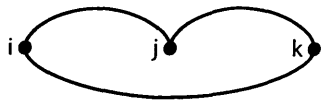

Figure 2-4: Diagram for the "loop" cumulant $\langle S_{ij}S_{jk}S_{ki}\rangle_c$.

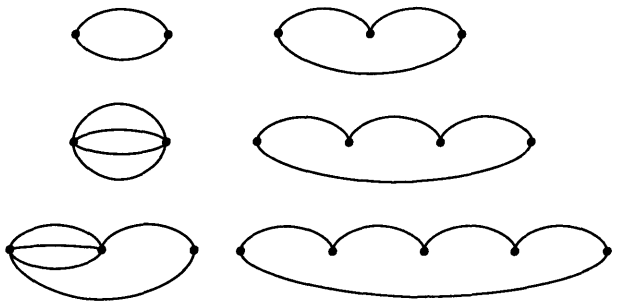

Figure 2-5: Nonvanishing cumulants up to order five.

$$\langle S_{ij}S_{kl}\rangle_c = \begin{cases} 1/4, & \text{if } ij, kl \text{ same pair} \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

For convenience we may represent higher order cumulants with diagrams. Fig. 2-4 is one example. Various rules may be applied to simplify calculations of these cumulants. Only one-point irreducible diagrams are nonvanishing, for example. Also, cumulants for topologically equivalent diagrams differ only by a factor related to the difference in the order (number of edges) of the two diagrams. The nonvanishing cumulants up to order five are depicted in Fig. 2-5.

From the means $\langle S_{ij}\rangle_c$ and covariances $\langle S_{ij}S_{kl}\rangle_c$ we conclude that the distribution of sequence points is centered at $(1/2, 1/2, \cdots, 1/2)$ and that the distribution is roughly rotationally symmetrical.

One-dimensional Projections of the Distribution

Another approach to analyzing the distribution of sequence points is to project them onto diagonals of the unit hypercube and to analyze the resulting one-dimensional distributions. For example, let I be the vector diagonally spanning the hypercube defined by $I_{ij} = 1$ for all i, j, and consider the distribution of the quantity $S \cdot I = S_{ij}I_{ij} = \sum_{i<j} S_{ij}$ over all sequence points S. This quantity ranges from a

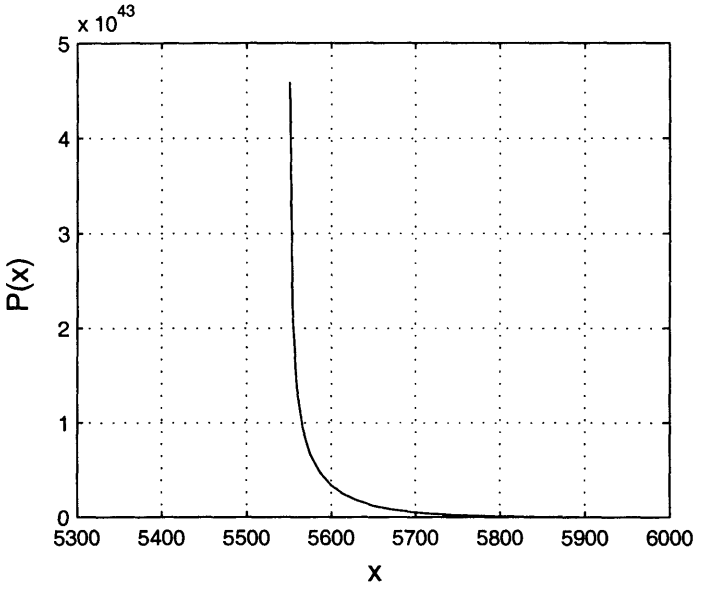

Figure 2-6: Distribution $P(x)$ of the quantity $x = S \cdot I$ over all sequence points S. Here $N = 150$ and $\langle x \rangle \approx 5590$.

minimum of $(N^2 - 2N)/4$ for sequence points whose corresponding sequences contain $N/2$ ones, i.e. sequences $\hat{S}$ for which $\sum_i \hat{s}_i = N/2$, to a maximum of $N(N-1)/2$ for sequence points whose corresponding sequences contain either all 1's or all 0's. A short calculation yields

$$P(x)dx \sim \frac{e^{-(4x-N^2+2N)/2N}}{\sqrt{4x - N^2 + 2N}}dx, \tag{2.4}$$

where $P(x)$ is a continuum approximation to the density of sequence points S for which $S \cdot I = x$. This distribution is highly skewed with respect to its mean $\langle x \rangle = N(N-1)/4$. See Fig. 2-6.

In fact, the distribution Eq. (2.4) holds for all 2^{N-1} diagonals of the hypercube that pass through a sequence point. See Appendix A. We expect the distribution to be similarly skewed for diagonals "near" one of these 2^{N-1} diagonals. However, due to the very high dimensionality of the contact space, there are still an overwhelming number of diagonals to account for. A calculation of the normalized cumulants of the one-dimensional projections of the sequence point distribution shows that, for the overwhelming majority of diagonals of the hypercube, this one-dimensional distribution approaches normality for large N. See Appendix A. We conclude that as a first

approximation we may take the distribution of sequence points to be a rotationally symmetric Gaussian centered at the center of the unit hypercube.

2.2.4 The Distribution of Structure Points

We now consider the distribution of RNA structure points F in contact space. For practical values of N, exhaustive enumeration of all structures is impossible. To investigate the distribution of structure points without full enumeration, we follow the procedure of calculating the moments $\langle F_{ij} \rangle$, $\langle F_{ij} F_{kl} \rangle$, etc. Taking the averages over all secondary structures of length N, we have $\langle F_{ij} \rangle = D_{N,ij}/X_N$, $\langle F_{ij} F_{kl} \rangle = D^{(2)}_{N,ijkl}/X_N$, etc., where $D^{(m)}_{N,ijkl\ldots}$ is the number of secondary structures of length N that contain the m base pairs (i,j), (k,l),.... These quantities may be recursively calculated. Explicitly, we have for the first two orders

$$D_{n+1,i,j} = D_{n,i,j} + \sum_{k=j+1}^{n-d} D_{k-1,i,j} X_{n-k} + \sum_{k=1}^{i-1} D_{n-k,i-k,j-k} X_{k-1} \tag{2.5}$$

$$\begin{aligned} D^{(2)}_{n+1,i,j,k,l} &= D^{(2)}_{n,i,j,k,l} + \sum_{p=1}^{i-1} X_{p-1} D^{(2)}_{n-p,i-p,j-p,k-p,l-p} \\ &\quad + \sum_{p=j+1}^{k-1} D_{p-1,i,j} D_{n-p,k-p,l-p} + \sum_{p=l+1}^{n-d} D^{(2)}_{p-1,i,j,k,l} X_{n-p}. \end{aligned} \tag{2.6}$$

The second recursion is for pairs (i,j) and (k,l) that are not nested, i.e. if $i < j$ and $k < l$, then either $i < j < k < l$ or $k < l < i < j$. If instead the pairs are nested, then a recursion similar to the recursion Eq. (2.5) for a single base pair applies.

A great simplification occurs if $d = 0$, i.e. if we ignore physical constraints and allow any two bases to form a base pair (including nearest-neighbors along the backbone). In this case we have

$$D_{n,i,j} = X_{j-i-1} X_{n-j+i-1} \quad (i < j) \tag{2.7}$$

$$D^{(2)}_{n,i,j,k,l} = \begin{cases} X_{j-i-1}X_{l-k-1}X_{n-j+i-l+k-2}, & i<j<k<l \\ X_{l-k-1}X_{j-i-l+k-2}X_{n-j+i-1}, & i<k<l<j \end{cases} . \tag{2.8}$$

In other words, to find the number of secondary structures $D_{n,i,j}$ with the bond (i,j), we simply multiply the number of structures possible for the interior region $i<x<j$ by the number of structures possible for the exterior region $(1 \le x < i) \cup (j < x \le n)$. The double-bond quantity $D^{(2)}_{n,i,j,k,l}$ similarly factorizes; the result depends on whether the bonds (i,j) and (k,l) are nested or not. This factorization clearly generalizes to the higher order quantities $D^{(m)}_{n,\ldots}$.

The means $\langle F_{ij} \rangle$ give the center of the distribution of structure points, while the covariances $\langle F_{ij}F_{kl} \rangle_c$ give the spreads in various directions. The matrix $C_{ijkl} = \langle F_{ij}F_{kl} \rangle_c$ is real and symmetric under $ij \leftrightarrow kl$; it can therefore be diagonalized to give the principal axes of the distribution along with the variances in those directions.

From the symmetry $C_{ijkl} = C_{N+1-i,N+1-j,N+1-k,N+1-l}$, we see that if F_{ij} is an eigenvector of C, then the vector $F_{N+1-i,N+1-j}$ is an eigenvector with the same eigenvalue. Thus if an eigenvalue is nondegenerate, its corresponding eigenvector possesses the mirror symmetry $F_{ij} = F_{N+1-i,N+1-j}$. If $d=0$, then in addition to mirror symmetry, C possesses the translational symmetry $C_{ijkl} = C_{i+p,j+p,k+p,l+p}$ for any integer p (taking the indices modulo N). We can utilize this symmetry by changing into Fourier variables with

$$\breve{F}_{qq'} = \frac{1}{N} \sum_{jj'} e^{i(qj+q'j')} (F_{jj'} - \langle F_{jj'} \rangle), \tag{2.9}$$

where q and q' range over $0, 2\pi/N, \ldots, 2\pi(N-1)/N$. Then the new matrix

$$\breve{C}_{qq'\bar{q}\bar{q}'} \equiv \langle \breve{F}_{qq'} \breve{F}^*_{\bar{q}\bar{q}'} \rangle = \frac{1}{N^2} \sum_{jj'\bar{j}\bar{j}'} e^{i(qj+q'j'-\bar{q}\bar{j}-\bar{q}'\bar{j}')} C_{jj'\bar{j}\bar{j}'} \tag{2.10}$$

is Hermitian and has the same eigenvalues as the original correlation matrix C_{ijkl}. Using translational symmetry to write $C_{ijkl} \to C_{y\bar{y}z}$, with $y \equiv j-i$, $\bar{y} \equiv l-k$, and $z \equiv k-i$, we find

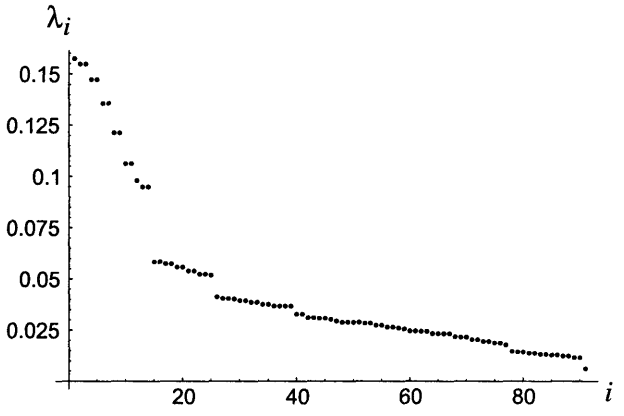

Figure 2-7: Eigenvalues of the correlation matrix C for $d = 0$ and $N = 14$. There are a total of 91 eigenvalues.

$$\breve{C}_{qq'\bar{q}\bar{q}'} = \frac{1}{N}\delta_{q+q'-\bar{q}-\bar{q}'(\mathrm{mod}\ 2\pi)} \sum_{y,\bar{y},z} e^{i(q'y-\bar{q}z-\bar{q}'(\bar{y}+z))} C_{y\bar{y}z} \tag{2.11}$$

so that the matrix elements $\breve{C}_{qq'\bar{q}\bar{q}'}$ vanish unless $q + q' = \bar{q} + \bar{q}'(\mathrm{mod}\ 2\pi)$. From this fact and the mirror symmetry $C_{ijkl} = C_{N+1-i,N+1-j,N+1-k,N+1-l}$ we conclude that the matrix $\breve{C}_{qq'\bar{q}\bar{q}'}$ is real, and in turn that $\breve{C}_{qq'\bar{q}\bar{q}'} = \breve{C}_{2\pi-q,2\pi-q',2\pi-\bar{q},2\pi-\bar{q}'}$.

We now analyze the spectrum of eigenvalues of $\breve{C}_{qq'\bar{q}\bar{q}'}$. Since the nonvanishing elements have $q + q' = \bar{q} + \bar{q}'$, the matrix is block diagonal, with blocks that may be labeled with their value of $q + q'$. The eigenvalues of the matrix are the union of the eigenvalues of the blocks. From the fact that $\breve{C}_{qq'\bar{q}\bar{q}'} = \breve{C}_{2\pi-q,2\pi-q',2\pi-\bar{q},2\pi-\bar{q}'}$, we see that the blocks $q + q' = n$ and $q + q' = -n(\mathrm{mod}\ 2\pi)$ have the same eigenvalues. Thus most eigenvalues come in pairs, the exceptions being the eigenvalues associated with the $q + q' = 0$ and $q + q' = \pi$ blocks (the latter occurring only when N is even). The eigenvectors associated with nondegenerate eigenvalues possess reflection symmetry, as noted earlier. Fig. 2-7 illustrates a typical set of eigenvalues.

Using the principal directions and variances given by the diagonalized correlation matrix, we can write down a Gaussian distribution that approximates the true structure point distribution in contact space. However, beyond having the correct first and second moments, the Gaussian may or may not resemble the true distribution. To investigate this numerically, we compared the approximate and true distributions

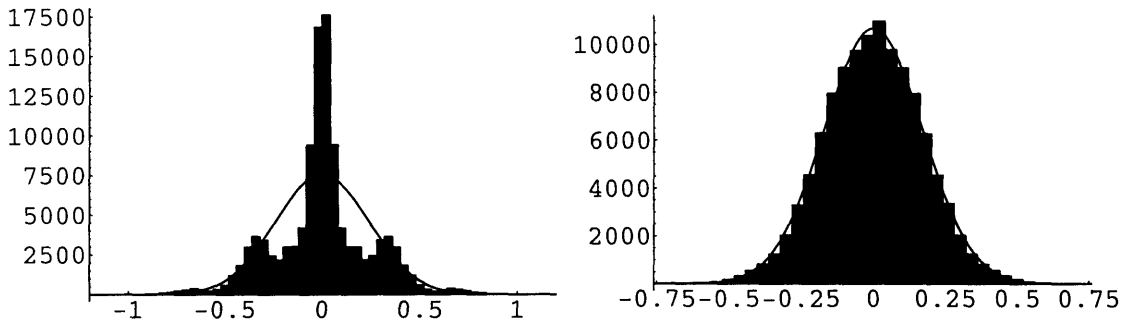

Figure 2-8: Comparison between the Gaussian approximation and the true structure point distribution for $d = 0$ and $N = 14$ in two principal directions. The histogram is the true structure point distribution, calculated through exhaustive enumeration of all structure points. Left: distributions projected along the 18th eigenvector, with eigenvalue $\lambda_{18} \approx 0.0575$. Right: distributions projected along the 50th eigenvector, with eigenvalue $\lambda_{50} \approx 0.0288$.

along the principal directions, using exhaustive enumeration to generate the true distribution. Fig. 2-8 shows some typical results. For the small values of N that are numerically accessible (say, up to $N \approx 16$), we find that the Gaussian approximation is appropriate for some principal directions but not others. Superficial observations indicate that the approximation works better for principal directions associated with smaller eigenvalues. Further investigation would be needed to determine why some directions in contact space posssess the non-normal distributions, and, if possible, the form of those distributions. It is possible that deviations from the Gaussian approximation disappear for larger values of N.

2.2.5 Discussion

Motivated by the work of Li et al. on the designability of protein structures [21], we have investigated the designability of RNA secondary structures within a simple toy model of RNA folding. In this model RNA bases come in two types and base pairing is favorable for complementary monomers. We use the contact representation for RNA structures and an analogous representation for RNA sequences so that both sequences and structures may be regarded as points in a single contact space. Insight into designability can be gained by investigating the distributions of sequence and

structure points in this space.

We find that one-dimensional projections of the sequence point distribution onto diagonals of the unit hypercube asymptotically approach the same normal distribution with increasing N. The interpretation of this result is not clear; it may be a simple consequence of the very high dimensionality of the contact space, as opposed to a meaningful characterization of the sequence point distribution. We performed a principal component analysis of the structure point distribution [34] for the ideal case $d = 0$, partially diagonalizing the correlation matrix analytically and finishing the diagonalization numerically. We compared the actual structure point distribution with a Gaussian distribution with the proper covariances by exhaustively enumerating all secondary structures of length N. For the very limited values of N that are accessible numerically, we find that the Gaussian approximation works in some principal directions but not others. Further investigation would be required to better understand this distribution.

2.3 Gapped RNA Secondary Structures

In this section we investigate the statistical mechanics of homopolymer RNAs, for which all the bases are equivalent. This highly simplified treatment directly describes physical RNAs only in the molten phase [4]. However, from a theoretical standpoint this toy model yields interesting results, such as a connection with ideal Gaussian polymers and a characterization of the length scale ξ beyond which RNAs are in the asymptotic or thermodynamic limit.

2.3.1 The Homopolymer RNA Partition Function Z_n

Recall that the number X_n of secondary structures of length n can be calculated recursively via the basic recursion Eq. (2.1). With the inclusion of a Boltzmann weight, this same recursion can be used to calculate the partition function Z_n for a homopolymer RNA of length n [4]:

$$Z_{n+1} = Z_n + \sum_{k=1}^{n-d} qZ_{k-1}Z_{n-k}; \qquad q = e^{-\beta\epsilon}. \tag{2.12}$$

Here ϵ is the energy of base pairing, and q is the corresponding Boltzmann weight.

We may use standard methods of combinatorics to derive the asymptotic behavior of the series Z_n (and therefore X_n, by setting $q = 1$) from Eq. (2.12). Multiplying both sides by z^n and applying $\sum_{n=0}^{\infty}$ leads to

$$\frac{1}{z}(\hat{Z} - 1) = \hat{Z} + qz\hat{Z}(\hat{Z} - \tilde{Z}_{d-1}), \tag{2.13}$$

where $\tilde{Z}_m \equiv 1 + z + z^2 + \cdots + z^m$ for $m \geq 0$, $\tilde{Z}_{-1} \equiv 0$, and $\hat{Z} = \hat{Z}(z) \equiv \sum_{n=0}^{\infty} Z_n z^n$ is the generating function for the series Z_n. Solving for $\hat{Z}$ gives

$$\hat{Z}(z) = \frac{1 - z + qz^2\tilde{Z}_{d-1} - \sqrt{(1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2}}{2qz^2}. \tag{2.14}$$

According to the formalism of *singularity analysis*, the singularities of $\hat{Z}(z)$ determine the asymptotic behavior of Z_n [11]. These singularities occur at the branch points of the square root in Eq. (2.14). The *dominant singularity* ρ is the singularity of smallest modulus, or the smallest root of the polynomial $\alpha(z) = (1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2$. $\alpha(z)$ and ρ depend on the minimum hairpin gap d; for $d = 0$ and $d = 1$, we have $\rho = 1/(1+2\sqrt{q})$ and $\rho = (1-\sqrt{1+4\sqrt{q}+2\sqrt{q}})/2q$, respectively. The large q behavior of $\rho(q)$ for $d \geq 1$ is given by[5]

$$\rho = \frac{1}{1+\sqrt{q}} - \frac{1}{q^{d/4+1/2}} + \cdots. \tag{2.15}$$

Fig. 2-9 shows the behavior of $\rho(q)$ as a function of q for various values of d.

Note that the leading large q behavior of ρ is different when $d = 0$ compared to when $d > 0$. As we saw in Sec. 2.2.4 with regards to translational invariance and the factorizability of the quantities $D_{n,i,j}$, $D^{(2)}_{n,i,j,k,l}$, etc., when $d = 0$ the system possesses many unique properties. In connection with the behavior of ρ at large q, we point out another of these properties. When $d > 0$, the ground state of the system is the unique

[5]See Appendix B.

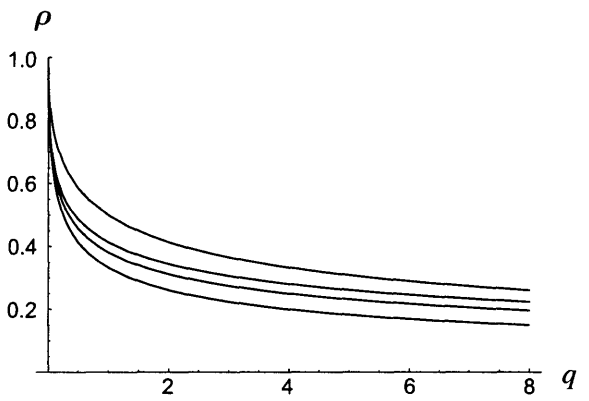

Figure 2-9: q dependence of the dominant singularity ρ for $d = 0, 1, 2, \infty$ (low to high).

Figure 2-10: Unique ground state secondary structure for $n = 12$, $d = 2$, and $q > 1$. In physical space this structure resembles a hairpin.

or weakly degenerate (depending on the parity of the number of bases n) "hairpin" structure depicted in Fig. 2-10. When $d = 0$ on the other hand, the ground state of the system is highly degenerate. It is plausible that the differing large q behavior of $\rho(q)$ when $d = 0$ as opposed to when $d > 0$ is a consequence of this difference in the degeneracy of the zero temperature state.

Returning to the asymptotic behavior of the series Z_n, we use tabulated results from singularity analysis to find [11]

$$Z_n = c_0 \frac{\rho^{-n}}{n^{3/2}} + c_1 \frac{\rho^{-n}}{n^{5/2}} + \cdots = c_0 \frac{\rho^{-n}}{n^{3/2}} \left[1 + \frac{c_1}{c_0 n} + \cdots \right] \tag{2.16}$$

for constants $c_0, c_1, \cdots$ that may be analytically calculated. The leading behavior $Z_n \sim \rho^{-n}/n^{3/2}$ is characteristic of the square-root singularity in the generating function $\hat{Z}(z)$. The ratio c_1/c_0 gives a crossover length scale for Z_n, or the length scale beyond which the RNA may be thought of as being in the asymptotic regime [22]. For $d = 0$, we have $c_1/c_0 = -9/4 - 3/(16\sqrt{q})$. In principle, we have a procedure for analytically calculating c_1/c_0 for any d; however, as d increases, exact calculations rapidly become too complicated to carry out. One thing we do find is that in the limit

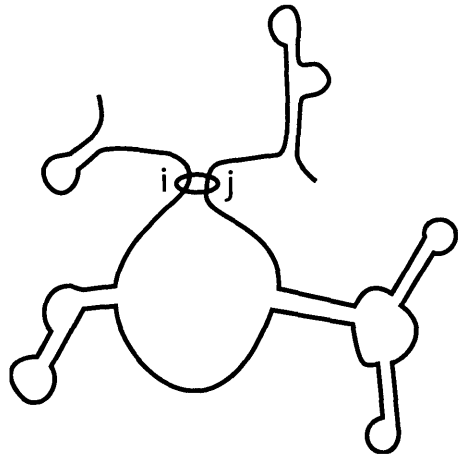

Figure 2-11: Slip-linked RNA.

of large q, we have $c_1/c_0 \sim q^{1/4}, q^{1/2}$ for $d = 1, 2$, respectively. Note the qualitative difference between the cases $d = 0$ and $d \neq 0$; for $d = 0$ the crossover scale does not increase with q.

2.3.2 The Restricted Partition Function $Z^{(r)}_{n,i,j}$ and Gapped Secondary Structures

We define the restricted partition function $Z^{(r)}_{n,i,j}$ to be the partition function for an RNA of length n in which the bases i and j are bonded. As observed previously in Sec. 2.2.4 in connection with the quantity $D_{n,i,j}$, the bond (i, j) serves to divide the RNA into two disjoint regions, the region interior to the bond and the region exterior to the bond. In the ideal case $d = 0$, we have the factorization analogous to Eq. (2.7):

$$Z^{(r)}_{n,i,j} = Z_{j-i-1} Z_{n-j+i-1}, \tag{2.17}$$

where the right-hand side involves unrestricted partition functions. Using the asymptotic value of Z_n gives

$$Z^{(r)}_{n,i,j} \sim \frac{\rho^{2-n}}{(j-i-1)^{3/2}(n-j+i-1)^{3/2}}. \tag{2.18}$$

This form for the restricted partition function leads to an interesting physical consequence. Suppose we imagine that an RNA is threaded through a *slip-link* [23].

Figure 2-12: Failure of simple factorization of $Z^{(r)}_{n,i,j}$ for $d \neq 0$. This secondary structure is not accounted for in the product $Z_{j-i-1}Z_{n-j+i-1}$ (assuming that one visualizes the factorization in the natural way).

See Fig. 2-11. The slip-link forces two bases i and j to bond, but the RNA is free to slide through the slip-link so that i and j may vary. In equilibrium, the RNA will explore the various possible configurations with weights given by $Z^{(r)}_{n,i,j}$. The particular form of $Z^{(r)}_{n,i,j}$ given by Eq. (2.18) dictates that the RNA will tend towards configurations where either the loop has length $\sim n^{1/2}$ and the exterior region has the remaining length $\sim n$, or vice versa. In other words, the loop in a slip-linked RNA will prefer to be either "tight" or "loose". This result is identical to a result obtained by Metzler et al. for slip-linked ideal Gaussian polymers [23]. The correspondence is the result of the factorizability of $Z^{(r)}_{n,i,j}$ along with the interesting coincidence that the asymptotic form $Z_n \sim \rho^{-n}/n^{3/2}$ is the same as the configurational entropy of an ideal polymer loop in three dimensions.

For the ideal case $d = 0$, the factorization in Eq. (2.17) generalizes to higher-order restricted partition functions in which additional pair bonds are enforced. The calculation of arbitrary restricted partition functions in this case is very simple and requires only knowledge of the behavior of Z_n. The consequences for more complicated slip-linked RNAs will be as given in [23].

We now consider the non-ideal cases $d > 0$. At first glance it may appear that the factorization of $Z^{(r)}_{n,i,j}$ still works; after all, the bond (i, j) still separates the RNA into an interior and an exterior region. The reason the factorization fails is that the partition function for the exterior region is no longer equal to $Z_{n-j+i-1}$. See Fig. 2-12.

To resolve this difficulty, we introduce the notion of a *gapped secondary structure.* A gapped secondary structure of length n, minimum hairpin gap d, and *gap* at p is a graph on n vertices with the usual properties, with the modification that base pairs (i, j) with $i \leq p < j$ do not have to obey the minimum gap requirement. In other

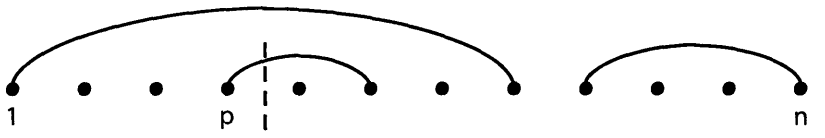

Figure 2-13: Gapped secondary structure with $d = 2$ and gap at p. Bonds that cross the dotted line do not have to obey the minimum gap requirement. Without the gap at p, this would not be a legal secondary structure.

words, such base pairs (i, j) may have $|j - i| \leq d$. See Fig. 2-13. We denote the partition function over gapped secondary structures with length n and gap at p by $Z^{(g)}_{n,p}$.

With this definition, the restricted partition function $Z^{(r)}_{n,i,j}$ again factors:

$$Z^{(r)}_{n,i,j} = Z_{j-i-1} Z^{(g)}_{n-j+i-1,i-1} \tag{2.19}$$

where the right-hand side now involves the partition function for secondary structures with gap at $i - 1$. Clearly, we may generalize the above definitions to doubly gapped, triply gapped, etc. secondary structures. Then, even if $d \neq 0$, restricted partition functions for RNAs with an arbitrary number of bonds will be factorizable into products of gapped and ungapped partition functions.

2.3.3 Analysis of the Gapped RNA Partition Function $Z^{(g)}_{n,p}$

The quantity $Z^{(g)}_{n,p}$ satisfies a recursion relation similar to the ones seen earlier:

$$Z^{(g)}_{n+1,p} = Z^{(g)}_{n,p} + q \sum_{k=1}^{p} Z_{k-1} Z^{(g)}_{n-k,p-k} + q \sum_{k=p+1}^{n-g} Z^{(g)}_{k-1,p} Z_{n-k}. \tag{2.20}$$

Multiplying both sides by $x^n y^p$ and applying $\sum_{n,p}^{\infty}$ gives

$$\hat{Z}^{(g)}(x, y) = \frac{\hat{Z}(xy)}{1 - x - qx^2 y \hat{Z}(xy) - qx^2[\hat{Z}(X) - \tilde{Z}_{d-1}(x)]}, \tag{2.21}$$

where $\hat{Z}^{(g)}(x, y) \equiv \sum_{n,p}^{\infty} Z^{(g)}_{n,p} x^n y^p$ is the bivariate generating function for the series $Z^{(g)}_{n,p}$ and as before $\tilde{Z}_m \equiv 1 + z + z^2 \cdots + z^m$. By performing asymptotics on $\hat{Z}^{(g)}(x, y)$ and its derivatives we can compute the asymptotic statistics of the series $Z^{(g)}_{n,p}$.

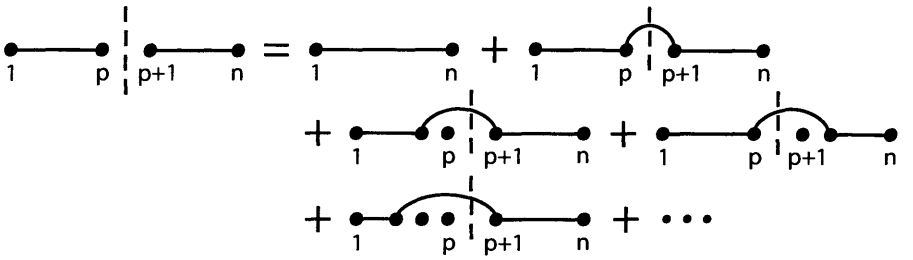

Figure 2-14: Diagrammatic representation of the recursion (2.22). Each diagram with a bond on the right-hand side represents the set of illegal diagrams with the depicted bond as the most interior bond relative to the gap at p.

There is an alternate recursion relation for $Z^{(g)}_{n,p}$ that will be useful. This relation stems from the observation that most gapped secondary structures on n bases are also ordinary secondary structures on n bases. The only gapped structures that do *not* qualify as ordinary secondary structures are those that have one or more bonds spanning the gap at p that do not meet the minimum hairpin requirement. We may classify these extra illegal diagrams by their innermost illegal bond, i.e. that bond which is most interior relative to the gap at p. The partition function for these sets of illegal diagrams can be written in terms of gapped partition functions. The net result is the recursion

$$\begin{aligned} Z^{(g)}_{n,p} &= Z_n + qZ^{(g)}_{n-2,p-1} + qZ^{(g)}_{n-3,p-2} + qZ^{(g)}_{n-3,p-1} \\ &\quad + \cdots + qZ^{(g)}_{n-d-1,p-d} + \cdots + qZ^{(g)}_{n-d-1,p-1} \\ &= Z_n + q\sum_{k=2}^{d+1}\sum_{l=1}^{k-1} Z^{(g)}_{n-k,p-l}. \end{aligned} \tag{2.22}$$

See Fig. 2-14 for a diagrammatic representation of this recursion. Note that the sum is truncated if $p < d$ or $p > n - d$.

For given n, the gapped partition function $Z^{(g)}_{n,p}$ can be viewed as a statistical distribution for p. The typical behavior of this distribution in the long RNA or large n limit is shown in Fig. 2-15. The gap p has roughly equal probability to be located anywhere along the RNA, except at the very ends where there is a drop-off in the distribution. Physically, this drop-off arises from entropic disadvantages that are

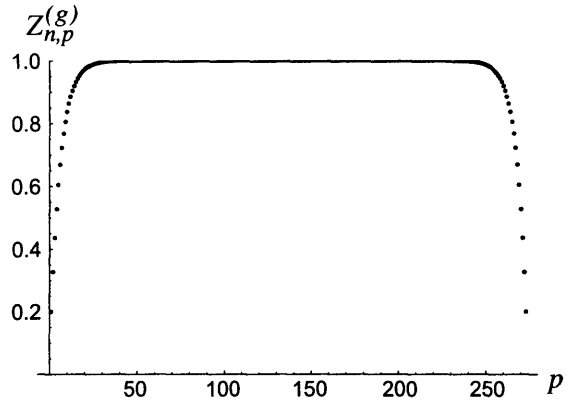

Figure 2-15: $Z^{(g)}_{n,p}$ (normalized) vs. p for a long RNA. Here $d = 3$, $q = 10$, and $n = 272$.

magnified (diminished) for $q > 1$ ($q < 1$). The value of $Z^{(g)}_{n,p}$ over the flat plateau can be computed using the recursion Eq. (2.22). Asymptotically (for large n) this quantity approaches the value[6] $Z_n/(1-r)$, where $r = q(\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1})$ and as before ρ is the dominant singularity for the polynomial $\alpha(z) = (1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2$. We also compute the length scale characterizing the edge effect; the drop-off at the tails of the distribution has an exponential form with a characteristic size that scales as $\xi \sim q^{d/4}$ for large q (see Appendix B).

The behavior of the gapped partition function provides an alternative length scale characterizing long RNAs, compared to the crossover scale c_1/c_0 discussed earlier in connection with Eq. (2.16): an RNA is long when $Z^{(g)}_{n,p}$ has the flat-top shape shown in Fig. 2-15. Intuitively we expect this length scale to be the same as that of the edge effect, with $\xi \sim q^{d/4}$, although the precise argument is not clear. An explicit approach to extracting this length scale is to follow the decay (with increasing n) of the normalized cumulants $\langle p^n \rangle_c / \langle p^2 \rangle_c^{n/2}$ to their asymptotic values for the large n distribution. In principle, these cumulants may be computed from the bivariate generating function Eq. (2.21); in practice the details are quite cumbersome. With some work using this approach we find that at large q the long RNA length scale behaves as $q^{1/4}$ and $q^{1/2}$ for $d = 1$ and $d = 2$, respectively, in agreement with the scale of the edge effect as expected. This also agrees with the scaling of the crossover length discussed earlier in connection with Eq. (2.16). It seems plausible that all

[6]See Appendix B.

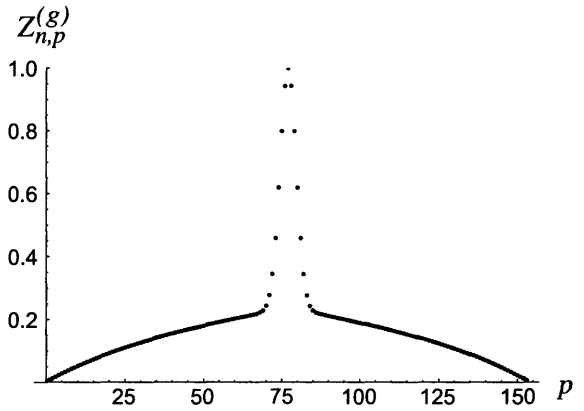

Figure 2-16: The distribution $Z_{n,p}^{(g)}$ (normalized) in the finite RNA regime. Here $d = 3$, $q = 30$, and $n = 152$.

these length scales are in fact the same, for all d.

When an RNA cannot be considered long, finite-size effects come into play. Fig. 2-16 shows a typical distribution $Z_{n,p}^{(g)}$ in this regime. The shape of the distribution is qualitatively different than that observed previously. As n is increased, the sharp central peak diminishes in relative size, becoming the flat plateau we saw earlier in the thermodynamic limit.

2.3.4 Discussion

We have explored a number of issues pertaining to the combinatorics and statistical mechanics of homopolymer RNA secondary structures. We began this section by calculating the asymptotic behavior of the homopolymer RNA partition function Z_n. A crossover length scale c_1/c_0 emerged characterizing the approach of Z_n to its leading asymptotic behavior for increasing n. We next investigated the restricted partition function $Z_{n,i,j}^{(r)}$ for RNAs with the bond (i, j). For the ideal case $d = 0$ this quantity factorizes, leading to an interesting connection with slip-linked ideal Gaussian polymers. The notion of the gapped secondary structure was introduced so that factorization of $Z_{n,i,j}^{(r)}$ would be possible for the non-ideal cases $d > 0$. We analyzed the behavior of the gapped partition function $Z_{n,p}^{(g)}$ in the long RNA limit and extracted a length scale characterizing when this limit is reached. To the extent that calculations could be made, this length scale agreed with the crossover scale

found for Z_n. We hypothesize that all the length scales are in agreement, behaving as $\xi \sim q^{d/4}$ for large q.

Appendix A

Results on the Distribution of Sequence Points

In this appendix we present detailed arguments for the results stated in Sec. 2.2.3. We begin by formalizing definitions.

Definitions An RNA *sequence* is a binary string $\hat{S} = \hat{s}_1\hat{s}_2\hat{s}_3\cdots\hat{s}_N$. We consider sequences with fixed length N. Let the sequences $\hat{S}$ exist in the N-dimensional *sequence space* U with Hamming metric $|\cdots|_U$, i.e. $|\hat{S}|_U = \sum_i |\hat{s}_i|$. We always denote sequence vectors with a hat.

For every sequence $\hat{S} \in U$, there is a corresponding *sequence point* $S \in V$ where V is the $\frac{N(N-1)}{2}$-dimensional *contact space*. A sequence point is constructed from its corresponding sequence by putting $S_{ij} = 1$ if $\hat{s}_i = \hat{s}_j$ and $S_{ij} = 0$ otherwise. Let the Hamming metric $|\cdots|_V$ be defined on the contact space.

The *inverse* A^* of a sequence point A is defined by $A^*_{ij} = 1$ if $A_{ij} = 0$ and $A^*_{ij} = 0$ if $A_{ij} = 1$. The inverse $\hat{S}^*$ of a sequence $\hat{S}$ is similarly defined. ∎

To analyze the distribution of sequence points in contact space we project the distribution onto diagonals of the unit hypercube. For diagonals that pass through a sequence point, this distribution is always the same.

Proposition 1 *Let $X \in V$ be a given sequence point and let X^* be its inverse. The points X and X^* define a diagonal $\tilde{X} = X^* - X$ of the unit hypercube in V.*

The distribution of the quantity $(S - X) \cdot \tilde{X}$ *over all sequence points* S *is the same, regardless of which sequence point* X *is considered.*

Proof Let S be a sequence point, and consider the sequences $\hat{X}$ and $\hat{S}$ corresponding to X and S. The distance $|S - X|_V$ is wholly determined by the distance $|\hat{S} - \hat{X}|_U$; if the latter is equal to n with $0 \leq n \leq N$, then the former is equal to $n(N-n)$. Noting that $\tilde{X} = -2X + I$ where $I_{ij} = 1$ for all i, j, we have $(S-X)\cdot\tilde{X} = (S-X)\cdot(-2X+I) = |S - X|_V$. Since the distribution of $n = |\hat{S} - \hat{X}|_U$ taken over all sequences $\hat{S}$ is independent of $\hat{X}$, the distribution of $|S - X|_V$ and hence $(S - X) \cdot \tilde{X}$ taken over all sequence points S is independent of X. ∎

The following proposition demonstrates that, as N goes to infinity, the projected one-dimensional distribution approaches a normal distribution for a fraction of all diagonals of the hypercube that approaches 1.

Proposition 2 *Let* C *be the set of vertices of the unit hypercube in* V. *For* $G \in C$, *define* $\tilde{G} = G^* - G$, *where* G^* *is the inverse of* G. *Then as* $N \to \infty$, *the distribution of the quantity* $(S - G) \cdot \tilde{G}$ *over all sequence points* S *approaches the same normal distribution for a fraction of the points* $G \in C$ *that approaches* 1.

Proof Let $G \in C$. First note that $\tilde{G} = -2G + I$ with $I = (1, 1, \cdots, 1)$ so that $\tilde{G}_{ij} = \left\{ \begin{array}{c} 1 \\ -1 \end{array} \right\}$ if $G_{ij} = \left\{ \begin{array}{c} 0 \\ 1 \end{array} \right\}$. Thus $(S-G)\cdot\tilde{G} = S\cdot\tilde{G}+|G|_V$ and $\langle[(S-G)\cdot\tilde{G}]^m\rangle_c = \langle(S \cdot \tilde{G})^m\rangle_c + \delta_{m,1}|G|_V$, where the averages are taken over all sequence points S.

We compute the cumulants $\langle(S \cdot \tilde{G})^m\rangle_c$. For the mean and variance we have

$$\begin{aligned}
\langle\tilde{G} \cdot S\rangle &= \tilde{G}_{ij}\langle S_{ij}\rangle && \text{(A.1)} \\
&= \frac{1}{2}\sum_{i<j}\tilde{G}_{ij} && \text{(A.2)} \\
&= M/2 - |G|_V && \text{(A.3)} \\
\langle(\tilde{G} \cdot S)^2\rangle_c &= \tilde{G}_{ij}\tilde{G}_{kl}\langle S_{ij}S_{kl}\rangle_c && \text{(A.4)} \\
&= \frac{1}{4}\sum_{i<j}(\tilde{G}_{ij})^2 = M/4 && \text{(A.5)}
\end{aligned}$$

where $M = \frac{N(N-1)}{2}$ and we have used Eqs. (2.2) and (2.3). Note that $\langle \tilde{G} \cdot S \rangle = M/2 - |G|_V \Rightarrow \langle (S - G) \cdot \tilde{G} \rangle = M/2$. These results are what we expect.

Now we continue to normalized higher order cumulants, and demonstrate that they tend to zero as $N \to \infty$. For third order cumulants we have

$$\begin{aligned} \langle (S \cdot \tilde{G})^3 \rangle_c &= \tilde{G}_{ij}\tilde{G}_{kl}\tilde{G}_{mn}\langle S_{ij}S_{kl}S_{mn}\rangle_c & \text{(A.6)} \\ &= \frac{1}{8}\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki}, & \text{(A.7)} \end{aligned}$$

because the only nonzero third-order cumulant $\langle S_{ij}S_{kl}S_{mn}\rangle_c$ is the "loop" cumulant $\langle S_{ij}S_{jk}S_{ki}\rangle = \frac{1}{8}$.

If G is a sequence point, then $\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki} = -1$ (no summation) always. The vast majority of points $G \in C$, however, are not sequence points. In fact, we have $\langle \tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki}\rangle_G = 0$, where the average is over all 2^M possible points $G \in C$. We also have

$$\langle (\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki})^2 \rangle_G = \langle \tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki}\tilde{G}_{mn}\tilde{G}_{no}\tilde{G}_{om}\rangle_G. \tag{A.8}$$

The only nonvanishing terms in the right-hand side occur when i, j, k and m, n, o are the same loop, i.e. when the two sets of indices are permutations of each other. Thus

$$\langle (\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki})^2 \rangle_G = (\text{combinatorial factor}) \times \begin{pmatrix} N \\ 3 \end{pmatrix} \sim N^3, \tag{A.9}$$

where the combinatorial factor in this case is just $3! \cdot 3!$. Thus the quantity $\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki}$ has 0 mean and standard deviation $\sim N^{3/2}$. If we consider the subset $Q \subset C$ of all points G for which $|\tilde{G}_{ij}\tilde{G}_{jk}\tilde{G}_{ki}| \leq N^2$, then as $N \to \infty$, Q encompasses the overwhelming majority of points $G \in C$. For these points, the normalized skewness behaves as

$$\frac{\langle (S \cdot \tilde{G})^3 \rangle_c}{\langle (S \cdot \tilde{G})^2 \rangle_c^{3/2}} \sim \frac{N^2}{N^3} \to 0. \tag{A.10}$$

The argument generalizes to the nth cumulant $\langle (S \cdot \tilde{G})^n \rangle_c$. Again, we have

$$\langle (S \cdot \tilde{G})^n \rangle_c = \tilde{G}_{ij} \tilde{G}_{kl} \cdots \langle S_{ij} S_{kl} \cdots \rangle_c. \tag{A.11}$$

In general, there are many different types or diagrams of nonvanishing cumulants contributing to the right-hand side. However, the "loop" diagram dominates over the other nonvanishing diagrams because there are far more possible loop diagrams than any other type. Thus

$$\langle (S \cdot \tilde{G})^n \rangle_c \sim \tilde{G}_{ij} \tilde{G}_{jk} \cdots \text{n-loop} \cdots. \tag{A.12}$$

Again, if G is a sequence point, then $\tilde{G}_{ij} \tilde{G}_{jk} \cdots = 1$ or -1 (no summation) always. When all points $G \in C$ are considered, however, we have $\langle \tilde{G}_{ij} \tilde{G}_{jk} \cdots \text{n-loop} \cdots \rangle_G = 0$ and

$$\begin{aligned} \langle (\tilde{G}_{ij} \tilde{G}_{jk} \cdots)^2 \rangle_G &= \langle \tilde{G}_{ij} \tilde{G}_{jk} \cdots \text{loop 1} \cdots \tilde{G}_{i'j'} \tilde{G}_{j'k'} \cdots \text{loop 2} \cdots \rangle_G & \text{(A.13)} \\ &\sim N^n. & \text{(A.14)} \end{aligned}$$

The quantity $\tilde{G}_{ij} \tilde{G}_{jk} \cdots \text{n-loop} \cdots$ has 0 mean and standard deviation $\sim N^{n/2}$. If we consider the subset $Q \subset C$ of points G for which $|\tilde{G}_{ij} \tilde{G}_{jk} \cdots| \leq N^{n/2+1}$, then as $N \to \infty$, Q contains the overwhelming majority of points $G \in C$. For these points, the relative nth cumulant $\frac{\langle (S \cdot \tilde{G})^n \rangle_c}{\langle (S \cdot \tilde{G})^2 \rangle_c^{n/2}} \sim \frac{N^{n/2+1}}{N^n} \to 0$. ∎

Appendix B

Results on RNA Combinatorics

In this appendix we give detailed arguments for some of the results presented in Sec. 2.3.

Lemma 1 *The large q behavior of the dominant singularity ρ of $\alpha(z) = (1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2$ for $d \geq 1$ is given by*

$$\rho = \frac{1}{1+\sqrt{q}} - \frac{1}{q^{d/4+1/2}} + \cdots. \tag{B.1}$$

Proof According to Pringsheim's theorem [11], the dominant root ρ is positive real if it is unique, which we assume. Using this fact we can rewrite the equation $\alpha(z) = 0$ as $\beta(x) = x^{d+2} - 2(1+\sqrt{q})x^{d+1} + (1+\sqrt{q})^2x^d - q = 0$ with $x = z^{-1}$. We look for the largest root w of this equation, with $w = \rho^{-1}$.

Numerically we find that the largest root is always near $x^* = 1 + \sqrt{q}$ for large q. Substituting $w = x^* + \delta$ in $\beta(w) = 0$ yields

$$\left(\frac{\delta}{x^*}\right)^2 \left(1 + \frac{\delta}{x^*}\right)^d = \frac{q}{x^{*d+2}}. \tag{B.2}$$

Thus to lowest order we have $(1+\sqrt{q})^d\delta^2 = q \Rightarrow \delta \approx q^{-d/4+1/2} \Rightarrow w = 1 + \sqrt{q} + q^{-d/4+1/2} + \cdots$ for large q. The lemma follows. ∎

Proposition 3 *Let a long RNA have length n. Let p be sufficiently far away from the edges 1 and n. Then asymptotically,*

$$Z_{n,p}^{(g)} = \frac{Z_n}{1-r}, \tag{B.3}$$

where $r = q(\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1})$ *and* ρ *is the dominant singularity for the polynomial* $\alpha(z) = (1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2$.

Before we prove the proposition we need a lemma.

Lemma 2 *The quantity* $r = q(\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1})$ *satisfies* $0 < r < 1$.

Proof of lemma That $r > 0$ is clear. To show that $r < 1$, we may rewrite the series $\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1}$ to yield

$$r = \frac{q(\rho^2 + \rho^3 + \cdots + \rho^{d+1}) - qd\rho^{d+2}}{1-\rho}. \tag{B.4}$$

Since ρ is a root of $\alpha(z) = (1 - z + qz^2\tilde{Z}_{d-1})^2 - 4qz^2$ and $\rho > 0$, we have $q(\rho^2 + \rho^3 + \cdots + \rho^{d+1}) = -1 + (1 + 2\sqrt{q})\rho$. Then

$$r = \frac{-1 + (1 + 2\sqrt{q})\rho - qd\rho^{d+2}}{1-\rho}. \tag{B.5}$$

Numerically we find that r approaches 1 when q is large. Using lemma 1 for the large q behavior of ρ gives

$$r = 1 - \frac{2}{q^{d/4}} + \cdots, \tag{B.6}$$

and $0 < r < 1$ as desired. ∎

Proof of proposition We begin with the recursion Eq. (2.22). By repeatedly substituting into the right-hand side, we can write $Z_{n,p}^{(g)}$ purely in terms of q and Z_m for various m:

$$Z_{n,p}^{(g)} = Z_n + q[Z_{n-2} + 2Z_{n-3} + \cdots + dZ_{n-d-1}] + q[q(Z_{n-4} + \cdots) + \cdots] + \cdots \tag{B.7}$$

Define a series $B_0, B_1, \cdots$ where B_i is the term in the above expansion involving q^i. Then $Z_{n,p}^{(g)} = B_0 + B_1 + B_2 + \cdots$. The series breaks down at some point, but if n is very large and p is sufficiently far away from the edges, then this breakdown does not occur for many terms. Using the asymptotic behavior of Z_n we have

$$\frac{B_2}{B_1} = \frac{q[Z_{n-2} + \cdots + dZ_{n-d-1}]}{Z_n} \tag{B.8}$$

$$= \frac{q\left[\frac{A\rho^{-(n-2)}}{(n-2)^{3/2}} + 2\frac{A\rho^{-(n-3)}}{(n-3)^{3/2}} + \cdots + d\frac{A\rho^{-(n-d-1)}}{(n-d-1)^{3/2}} + O\left(\frac{\rho^{-(n-2)}}{n^{5/2}}\right)\right]}{A\frac{\rho^{-n}}{n^{3/2}}} \tag{B.9}$$

$$= q\left[\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1} + O\left(\frac{1}{n}\right)\right] \tag{B.10}$$

$$\approx r. \tag{B.11}$$

In fact, to the extent that n is large and p is far away from the edges, this ratio holds for any two consecutive terms B_j and B_{j+1}, i.e. $\frac{B_{j+1}}{B_j} \to r$. Thus we have

$$Z_{n,p}^{(g)} = B_0 + B_1 + B_2 + \cdots \tag{B.12}$$

$$= Z_n + rZ_n + r^2 Z_n + \cdots. \tag{B.13}$$

According to the lemma, the series converges, and the result follows. ∎

Proposition 4 *The edge effect in the probability distribution $Z_{n,p}^{(g)}$ for the gap location p has the form of an exponential with characteristic decay length $\xi \sim q^{d/4}$ for large q.*

Proof For $p < n/2$ we propose the form

$$Z_{n,p}^{(g)} = \frac{Z_n}{1-r}\left[1 - e^{-p/\xi}\right], \tag{B.14}$$

where as before $r = q(\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1})$. The factor $Z_n/(1-r)$ is just the "bulk" value of $Z_{n,p}^{(g)}$ away from the edge $p = 1$.

To find the decay length ξ, we substitute (B.14) into the recursion Eq. (2.22). In doing so, we implicitly assume that $p > d$, since otherwise the recursion Eq. (2.22)

does not hold. Thus our argument is relevant only when q is so large that $\xi > d$, or in other words when there is a relatively slow or continuous decay of $Z_{n,p}^{(g)}$ near the edge $p = 1$. Substitution yields

$$q(\rho^2 + \rho^3 + \cdots + \rho^{d+1})e^{1/\xi} + q(\rho^3 + \cdots + \rho^{d+1})e^{2/\xi} + \cdots + q\rho^{d+1}e^{d/\xi} = 1. \quad \text{(B.15)}$$

This equation may be solved to find ξ. When q is large so that $1/\xi$ is small, we may expand the exponentials and

$$q\rho^2\left[1 + \frac{1}{\xi}\right] + q\rho^3\left[2 + \frac{3}{\xi}\right] + \cdots + q\rho^{d+1}\left[d + \frac{d(d+1)/2}{\xi}\right] = 1 \quad \text{(B.16)}$$

For large q we expect $\xi \sim q^m$ with $m > 0$; keeping only terms of order $q\rho^2/\xi$ or larger gives

$$q\rho^2\frac{1}{\xi} + q(\rho^2 + 2\rho^3 + \cdots + d\rho^{d+1}) = 1 \quad \text{(B.17)}$$

$$q\rho^2\frac{1}{\xi} + r = 1. \quad \text{(B.18)}$$

From lemma 1 we have $r = 1 - 2/q^{d/4} + \cdots$, so that $\xi \sim q^{d/4}$. ∎

Bibliography

[1] See, for example, the movies from the Theriot group at http://cmgm.stanford.edu/theriot/movies.htm.

[2] See http://www.nag.com/numeric/fl/manual/html/FLlibrarymanual.asp.

[3] G. Ahlers, M. C. Cross, P. C. Hohenberg, and S. Safran. The amplitude equation near the convective threshold: application to time-dependent heating experiments. *J. Fluid Mech.*, 110:297–334, 1981.

[4] R. Bundschuh and T. Hwa. Rna secondary structure formation: A solvable model of heteropolymer folding. *Phys. Rev. Lett.*, 83:1479–1482, 1999.

[5] L. Cameron, M. J. Footer, A. van Oudenaarden, and J. A. Theriot. Motility of acta protein-coated microspheres driven by actin polymerization. *Proc. Natl. Acad. Sci. U.S.A.*, 96:4908–4913, 1999.

[6] L. A. Cameron, P. A. Giardini, F. S. Soo, and J. A. Theriot. Secrets of actin-based motility revealed by a bacterial pathogen. *Nat. Rev. Mol. Cell Biol.*, 1:110–119, 2000.

[7] L. A. Cameron, J. R. Robbins, M. J. Footer, and J. A. Theriot. Biophysical parameters influence actin-based movement, trajectory, and initiation in a cell-free system. *Mol. Biol. Cell*, 15:2312–2323, 2004.

[8] M. F. Carlier, C. L. Clainche, S. Wiesner, and D. Pantaloni. Actin-based motility: from molecules to movement. *BioEssays*, 25:336–345, 2003.

[9] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65:851–1112, 1993.

[10] G. A. Dabiri, J. M. Sanger, D. A. Portnoy, and F. S. Southwick. *Listeria monocytogenes* moves rapidly through the host-cell cytoplasm by inducing directional actin assembly. *Proc. Natl. Acad. Sci. U.S.A.*, 87:6068–6072, 1990.

[11] P. Flajolet and R. Sedgewick. Notes from *The Average Case Analysis of Algorithms*, see http://algo.inria.fr/flajolet/Publications/books.html.

[12] A. B. Groswasser, S. Wiesner, R. M. Golsteyn, M. F. Carlier, and C. Sykes. The dynamics of actin-based motility depend on surface parameters. *Nature*, 417:308–311, 2002.

[13] J. Gucht, E. Paluch, J. Plastino, and C. Sykes. Stress release drives symmetry breaking for actin-based movement. *Proc. Natl. Acad. Sci. U.S.A.*, 102:7847–7852, 2005.

[14] H. Haken. *Synergetics.* Springer–Verlag, Berlin, 1983.

[15] J. Happel and H. Brenner. *Low Reynolds Number Hydrodynamics.* Prentice–Hall, 1965.

[16] P. G. Higgs. Rna secondary structure: physical and computational aspects. *Quarterly Reviews of Biophysics*, 33:199–253, 2000.

[17] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of rna secondary structures. *Monatshefte fur Chemie*, 125:167–188, 1994.

[18] P. S. Hubbard. Rotational brownian motion. *Phys. Rev. A*, 6:2421–2433, 1972.

[19] C. Kocks, R. Hellio, P. Gounon, H. Ohayon, and P. Cossart. Polarized distribution of *Listeria monocytogenes* surface protein acta at the site of directional actin assembly. *J. Cell Sci.*, 105:699–710, 1993.

[20] H. Li, R. Helling, C. Tang, and N. Wingreen. Emergence of preferred structures in a simple model of protein folding. *Science*, 273:666–669, 1996.

[21] H. Li, C. Tang, and N. S. Wingreen. Are protein folds atypical? *Proc. Natl. Acad. Sci. U.S.A.*, 95:4987–4990, 1998.

[22] T. Liu and R. Bundschuh. Analytical description of finite size effects for rna secondary structures. *Phys. Rev. E*, 69:061912, 2004.

[23] R. Metzler, A. Hanke, P. G. Dommersnes, Y. Kantor, and M. Kardar. Tightness of slip-linked polymer chains. *Phys. Rev. E*, 65:061103, 2002.

[24] A. Mogilner and G. Oster. Force generation by actin polymerization ii: The elastic ratchet and tethered filaments. *Biophys. J.*, 84:1591–1605, 2003.

[25] V. Noireaux, R. M. Golsteyn, E. Friederich, J. Prost, C. Antony, D. Louvard, and C. Sykes. Growing an actin gel on spherical surfaces. *Biophys. J.*, 78:1643–1654, 2000.

[26] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1992.

[27] H. Risken. *The Fokker-Planck Equation: Methods of Solutions and Applications.* Springer–Verlag, New York, 1996.

[28] S. Samarin, S. Romero, C. Kocks, D. Didry, D. Pantaloni, and M. F. Carlier. How vasp enhances actin-based motility. *J. Cell. Biol.*, 163:131–142, 2003.

[29] I. M. Schwartz, M. Ehrenberg, M. Bindschadler, and J. L. McGrath. The role of substrate curvature in actin-based pushing forces. *Curr. Biol.*, 14:1094–1098, 2004.

[30] K. Sekimoto, J. Prost, F. J'ulicher, H. Boukellal, and A. Bernheim-Grosswasser. Role of tensile stress in actin gels and a symmetry-breaking instability. *Eur. Phys. J. E*, 13:247–259, 2004.

[31] S. Strogatz. *Nonlinear Dynamics and Chaos.* Addison Wesley, Reading, MA, 1994.

[32] A. van Oudenaarden and J. A. Theriot. Cooperative symmetry-breaking by actin polymerization in a model for motility. *Nature Cell Biol.*, 1:493–499, 1999.

[33] S. Wiesner, E. Helfer, D. Didry, G. Ducouret, F. Lafuma, M. F. Carlier, and D. Pantaloni. A biomimetic motility assay provides insight into the mechanism of actin-based motility. *J. Cell. Biol.*, 160:387–398, 2003.

[34] M. Yahyanejad, M. Kardar, and C. Tang. Structure space of model proteins: A principal component analysis. *J. Chem. Phys.*, 118:4277–4284, 2003.

[35] D. Yarar, W. To, A. Abo, and M. D. Welch. The wiskott-aldrich syndrome protein directs actin-based motility by stimulating actin nucleation with the arp2/3 complex. *Curr. Biol.*, 9:555–558, 1999.