

Exact Folding Dynamics of RNA Secondary Structures

Michael T. Wolfinger^a, W. Andreas Svrcek-Seiler^a, Christoph Flamm^a, Ivo L. Hofacker^{a§}, and Peter F. Stadler^b

^aInstitut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria

^bBioinformatik, Institut für Informatik, Universität Leipzig, D-04103 Leipzig, Germany

Abstract.

Barrier trees consisting of local minima and their connecting saddle points imply a natural coarse-graining for the description of the energy landscape of RNA secondary structures. Here we show that, based on this approach, it is possible to predict the folding behavior of RNA molecules by numerical integration. Comparison with stochastic folding simulations shows reasonable agreement of the resulting folding dynamics and a drastic increase in computational efficiency that makes it possible to investigate the folding dynamics of RNA of at least tRNA size. Our approach is readily applicable to bistable RNA molecules and promises to facilitate studies on the dynamic behavior of RNA switches.

§ To whom correspondence should be addressed.

Email: ivo@tbi.univie.ac.at, Phone: **43 1 4277 52736, Fax: **43 1 4277 52793

1. Introduction

The comprehensive understanding of the folding process of biopolymers such as proteins and nucleic acids is one of the core issues in structural biology. It seems fair to say that Molecular Mechanics, despite all the progress in recent years [6, 33], will for the foreseeable future remain incapable of predicting, say, the folding pathway of a globular protein starting from a random coil state all the way to its (unknown) native state. Obviously, such a simulation would solve the protein folding problem.

In contrast to protein folding, the secondary structures of nucleic acids provide a level of description that is sufficient to understand the thermodynamics and kinetics of RNA folding [8]. We show here that folding pathways of RNA molecules at least the size of tRNAs can be computed for arbitrarily long time-scales within the secondary structure framework. To this end we exploit the fact that RNA secondary structures can be computed exactly with efficient polynomial algorithms, a fact that allows a detailed computational analysis of the conformational energy landscape.

A landscape perspective was used with some success also in theoretical work on protein folding [11, 27, 29], but a direct computational analysis of the landscapes of real molecules is possible only for very short peptides, see e.g. the review of Klepeis [18]. A master-equation approach similar to ours has been reported in [28] for lattice proteins, but in spite of the drastic simplification of the lattice model it is still computationally harder than the RNA case.

A thorough analysis of RNA folding dynamics is a necessary prerequisite toward understanding the functionality of a variety of small RNA molecules. It has been shown repeatedly that alternative conformations of the same RNA sequence can perform completely different functions, e.g. [2, 30, 37]. SV11, for instance, is a relatively small molecule that is replicated by Q_β replicase. It exists in two major conformations, a meta-stable multi-component structure and a rod-like conformation, constituting the native state, separated by a huge energy barrier. While the meta-stable conformation is a template for Q_β replicase, the ground state is not. By melting and rapid quenching the molecule can be re-converted from the inactive stable to the active meta-stable form [45].

In recent years dynamical aspects of RNA structure formation, including transitions at the level of RNA secondary structure, have received increasing attention, because they can play a crucial role for the understanding of the biological function of RNA. It has been shown for a number of natural RNAs that the formation of alternative or metastable conformations are well-defined steps in their folding pathways. These folding intermediates determine the biological function of the molecule.

The translation of the four genes encoded on the genomic RNA of the bacteriophage MS2 is regulated by the secondary structure transition of the 5' untranslated leader sequence from a metastable hairpin to a stable cloverleaf structure [34]. While the expression of the lysis and replicase genes is coupled to the expression of the coat protein in the full-length RNA, the maturation gene, coding for the A-protein needed by the

virion for the attachment to *E. coli*, is inaccessible to the ribosome due to the cloverleaf structure of the leader sequence. During transcription of the viral RNA the 5'-end of the leader sequence is trapped in a metastable hairpin allowing the ribosome to access the A-protein gene. After some time the hairpin is disrupted in favor of the stable cloverleaf, thereby silencing the A-protein gene expression. This secondary structure switch precisely controls the amount of A-protein translated from the MS2 genomic RNA.

The Hok/Sok system of plasmid R1 from *E. coli* is another prominent example for the regulation of gene expression via an intricate cascade of secondary structural rearrangements. The Hok/Sok system mediates plasmid maintenance by expressing the Hok toxin which kills plasmid-free segregates. The plasmid encodes for a highly stable mRNA, which is translated to the Hok toxin if the mRNA is in its activated conformation, and a labile anti-sense RNA (Sok) which act as an antidote by binding to the activated *hok* mRNA, leading to rapid degradation of the resulting duplex. The full-length *hok* mRNA forms a pool of inactive mRNAs. In time, however, the *hok* mRNA gets processed resulting in the truncation of the 3'-end, which triggers a refolding of the mRNA into the active conformation. Then both locations, the Hok gene and the Sok binding site are accessible. If the plasmid was lost, the pool of the antidote Sok is depleted, since the *hok* mRNA is considerably more stable than the *sok* RNA inducing the killing of the cell. For recent reviews on biologically functional RNA switches we refer to [4] and [24].

2. The Energy Landscape of RNA Molecules

RNA secondary structures can be decomposed uniquely into a set of “loops” of different types: stacked base pairs, bulges, interior loops, and multi-branched loops. The standard energy model [21] describes the energy of an RNA secondary structure as a sum of sequence-dependent contributions for each loop. Dynamic programming algorithms are known to exactly and efficiently compute the minimum free energy structure [48], the base pairing probability matrix [22], the density of states [7], certain sub-optimal structures [47], or all structures with an energy below a threshold value [44]. A suite of these algorithms is implemented in the **Vienna RNA Package** which forms the basis for the computations reported here [14, 15].

Within the framework of RNA secondary structures we can understand the process of folding as a time-series of secondary structures such that the elementary transitions are the opening or closing of a single base pair. This idea is implemented in the program **kinfold** [9], which allows simulations of RNA folding trajectories for macroscopic time scales. The simulated annealing approach to secondary structure formation used in [36] is based on the same idea. Other methods to simulate folding dynamics typically use formation and deletion of helices as the move set [13, 20, 23], but this requires ad hoc assumptions about the rates. The **paRNass** [40] program tries to predict RNA switches by clustering suboptimal structures by structural similarity and a crude measure of

the energy barrier between the clusters. Here we undertake a much more detailed investigation of the energy landscape.

Given an RNA sequence s , let X denote the set of possible secondary structures that can be formed from s satisfying the pairing logic of RNA, i.e., considering only Watson-Crick (GC, AU) and wobble (GU) pairs. The move-set \mathcal{M} (e.g. opening and closing of base pairs) and the energy function E define a *landscape* on X that can be seen as a coarse grained (discrete) version of the potential energy surfaces used e.g. in the MD simulations.

Within the framework of the folding landscape we can meaningfully speak of local minima or metastable states, their basins of attraction, and the saddle points separating them. Formally, a secondary structure $x \in X$ is a local minimum of E if $E(x) \leq E(y)$ for all its neighbors, $(x, y) \in \mathcal{M}$. A gradient walk is defined as follows: starting from $x \in X$ we move to its neighbor y with minimal energy if $E(y) < E(x)$. If the minimum energy neighbor y of x is not uniquely defined we use a deterministic rule to break the tie, for instance, by choosing the structure that comes lexicographically first. The step from x to $y = \gamma(x)$ is repeated until we reach a local minimum where the walk terminates, $\gamma(x) = x$. The local minima are therefore the attractors of the map $\gamma : X \rightarrow X$ and each $x \in X$ is mapped to a unique local minimum $z = \gamma^\infty(x) = \gamma^t(x)$ by a finite number t of applications of γ . The basin of attraction of a local minimum z , $\mathcal{B}(z)$, consists of all secondary structures that are mapped to it by the gradient walk, i.e. $\mathcal{B}(z) = \{x \in X | \gamma^\infty(x) = z\}$. Below we will need the (trivial) fact that these “gradient basins” of the local minima form a partition of X .

Let us now turn to the transitions between local minima. The energy of the lowest saddle point separating two local minima x and y is

$$E[x, y] = \min_{\mathbf{p} \in \mathbb{P}_{xy}} \max_{z \in \mathbf{p}} E(z) \quad (1)$$

where \mathbb{P}_{xy} is the set of all paths \mathbf{p} connecting x and y by a series of subsequent moves. The saddle-point energy $E[., .]$ is an ultra-metric distance measure on the set of local minima, see e.g. [35].

In the simplest case the energy function is non-degenerate, i.e., $f(x) = f(y)$ implies $x = y$. Then there is a unique saddle point $s = s(x, y)$ connecting x and y characterized by $E(s) = E[x, y]$. This definition of a saddle point is more restrictive than in differential geometry where saddles are not required to separate local optima [39]. For each saddle point s there exists a unique collection of configurations $\mathcal{V}(s)$ that can be reached from s by a path along which the energy never exceeds $E(s)$. In other words, the configurations in $\mathcal{V}(s)$ are mutually connected by paths that never go higher than $E(s)$. This property warrants to call $\mathcal{V}(s)$ the *valley below the saddle* s . Furthermore, suppose that $E(s) < E(s')$. Then there are two possibilities: if $s \in \mathcal{V}(s')$ then $\mathcal{V}(s) \subseteq \mathcal{V}(s')$, i.e., the valley of s is a “sub-valley” of $\mathcal{V}(s')$, or $s \notin \mathcal{V}(s')$ in which case $\mathcal{V}(s) \cap \mathcal{V}(s') = \emptyset$, i.e., the valleys are disjoint. This property arranges the local minima and the saddle points in a unique hierarchical structure which is conveniently represented as a tree, termed *barrier tree* (see Fig. 1a). Since saddle points separate local optima, each valley $\mathcal{V}(s)$

contains (in the non-degenerate case at least two) local minima z_1, \dots, z_k . Conversely, $\mathcal{V}(s) \subseteq \bigcup_k \mathcal{B}(z_k)$, i.e., the valley of s is contained in the union of the basins of attraction of the metastable states “below” s . The metastable states therefore form the tips (or leafs) of the barrier tree. In the case of degenerate landscapes an analogous construction is possible when certain saddle points with the same energy are collected into equivalence classes. For the mathematical details we refer to [10].

The exact calculation of the barrier tree for *discrete* systems is a highly challenging computational problem and only recently some progress in that direction has been achieved. Even for very small system sizes a simple-minded exhaustive search approach to evaluating Eq. (1) would be hopeless as one must calculate all paths connecting all pairs of minima. In fact, the exact evaluation of these paths was qualified as “desperately hard” in a bold earlier study of the energy barriers of the SK model [25], see also [19]. Instead, the program package `barriers`|| constructs the barrier tree directly from an energy sorted list of all configurations [9]. Starting with the lowest energy configurations, `barriers` explicitly builds the valleys $\mathcal{V}(s)$ and subtrees by checking for each configuration whether it (a) is a local minimum, (b) uniquely belongs to the basin of a local minimum that was encountered earlier in the list, or (c) “merges” two or more basins, i.e., whether it is a saddle point. In contrast, methods for exploring the *continuous* energy surfaces of molecules and molecular clusters, for which exact enumeration is impossible, use incomplete databases of minima, transition states and their connecting rearrangements. When searching transition states starting from a given minimum, these numerical techniques make explicit use of the fact that the potential energy surface is a differentiable manifold and hence are not applicable to the discrete setting considered here; see e.g. [1, 3, 41, 42].

In practice we usually cannot generate the complete landscape of the RNA molecule. The program `RNAsubopt` [44], however, computes all structures below a certain threshold value in $\mathcal{O}(n^3 + nQ)$ time, where n is the sequence length and Q is the number of secondary structures of interest. This is a controlled approximation in the sense that we can compare the partition function of all explicitly generated structures x_i , $Z' = \sum_{i=1}^Q \exp(-E(x_i)/RT)$, with the exact partition function Z computed by McCaskill’s algorithm: $(Z - Z')/Z$ is the total equilibrium frequency of the omitted high energy structures. In practice it is sufficient to use an energy band that extends only a few kT above the highest energy structure of interest, such that $(Z - Z') \ll Z$.

3. Kinetics

Very few experimental data are available at present to estimate transition rates between different secondary structures. It is known that the rate of hairpin formation is governed by the cancellation of the positive loop energy by the closing base-pair [31, 32], and that local hairpin formation is favored over long-distance structural elements, because of the spatial proximity of the opposing base-pair partners [5, 26].

|| The software is available from <http://www.tbi.univie.ac.at/~ivo/RNA/Barriers/>.

In [9] it has been shown that a good approximation to the few available quantitative and qualitative data on RNA folding kinetics is obtained by modeling conformational changes in terms of elementary steps of opening and closing of base pairs. In this approach the transition rate r_{xy} from secondary structure y to secondary structure x is non-zero only if $(y, x) \in \mathcal{M}$, i.e., if x and y are neighbors in the conformational energy landscape. In other words, the folding dynamics of an RNA molecule is described as motion on its conformational energy surface. Denote the probability that the molecule has the secondary structure x at time t by $p_x(t)$, the dynamics is governed by the master equation

$$\frac{dp_x}{dt} = \sum_{y \in X} r_{xy} p_y(t), \quad \text{with } r_{xx} = - \sum_{y \neq x} r_{yx}. \quad (2)$$

In other words, the dynamics is described by a continuous-time Markov process with infinitesimal generator $\mathbf{R} = (r_{yx})$. We solve this linear system of differential equations by explicitly computing $\vec{p}(t) = \exp(t\mathbf{R})\vec{p}(0)$.

The only missing ingredient is a model for the rates r_{xy} between *neighboring* RNA secondary structures. The transition state model dictates an expression of the form

$$r_{yx} = r_0 e^{-\frac{E_{yx}^\ddagger - E(x)}{RT}} \quad \text{for } x \neq y \quad (3)$$

where the transition state energies E_{yx}^\ddagger must be symmetric to assure detailed balance, $E_{yx}^\ddagger = E_{xy}^\ddagger$. In the simplest case one can use

$$E_{yx}^\ddagger = \max\{E(x), E(y)\} \quad (4)$$

which amounts to the Metropolis rule of simulated annealing. More sophisticated models for RNA are discussed in [17, 36, 38]. The parameter r_0 could be used to gauge the time axis from experimental data, here we simply use $r_0 = 1$.

4. Macro-States and Transition Rates

A description of the energy landscape or the dynamics of an RNA molecule based on all secondary structures is feasible only for very small sequences since the number of structures $|X|$ grows exponentially with sequence length [16, 43]. We therefore need to coarse-grain the representation of the energy landscape.

The simplest and most straightforward approximation for the folding dynamics is the Arrhenius law for transitions on the barrier tree. Within this model, transitions occur only between local minima that are directly connected by a saddle point, and the transition state energies are approximated by the saddle point energy $E[\alpha, \beta]$. This approximation completely neglects entropic terms that arise because there are many possible paths connecting two local minima.

A much better approximation can be derived from the microscopic dynamics as follows. Let $\mathbf{\Pi} = \{\alpha, \beta, \dots\}$ be a partition of the state space X . The classes of such a partition are *macro-states*. As a concrete example consider the partition of X defined

by the gradient basins $\mathcal{B}(z)$ of the local energy minima. To each macrostate α we can assign the partition function

$$Z_\alpha = \sum_{x \in \alpha} e^{-E(x)/RT} \quad (5)$$

and the corresponding free energy

$$G(\alpha) = -RT \ln Z_\alpha \quad (6)$$

Let us now turn to the transitions between macrostates. Suppose we know the transition rates r_{yx} from x to y . Then

$$r_{\beta\alpha} = \sum_{y \in \beta} \sum_{x \in \alpha} r_{yx} \text{Prob}[x|\alpha] \quad \text{for } \alpha \neq \beta \quad (7)$$

where $\text{Prob}[x|\alpha]$ is the probability to occupy state $x \in \alpha$ given that we know that the process is in macrostate α . The kinetics of the molecule in terms of its macro-states is given by the master equation

$$\frac{dp_\alpha}{dt} = \sum_{\beta \in \Pi} r_{\alpha\beta} p_\beta(t) \quad (8)$$

where $p_\alpha(t) = \sum_{x \in \alpha} p_x(t)$ and $r_{\alpha\alpha} = -\sum_{\beta \neq \alpha} r_{\alpha\beta}$. Assuming (local) equilibrium we have $\text{Prob}[x|\alpha] = e^{-E(x)/RT} / Z_\alpha$ and hence

$$r_{\beta\alpha} = \frac{1}{Z_\alpha} \sum_{y \in \beta} \sum_{x \in \alpha} r_{yx} e^{-E(x)/RT} \quad (9)$$

The point here is that we can compute $r_{\beta\alpha}$ “on flight” while executing the **barriers** program if two conditions are satisfied: (a) For each x we can efficiently determine to which macro-state it belongs and (b) the double sum in Eq. (9) needs to be evaluated only for neighboring conformations $(x, y) \in \mathcal{M}$. Condition (b) is obviously satisfied in the landscape model since $r_{yx} = 0$ by definition unless x and y are neighbors.

Condition (a) is easily satisfied for each of the gradient basins: in each step of the **barriers** algorithm all neighbors y of the newly added structures x that have a smaller energy have already been processed. Hence, if their assignment to a gradient basin is known, the assignment for x equals the one for its lowest energy neighbor. Initially, each local optimum forms the nucleus of new gradient basin, hence the macro-state to which x belongs can be determined in $\mathcal{O}(\delta)$ operations, where δ is the maximum number of neighbors of a secondary structure.

We can use the transition state model to define the free energies of the transition state $G_{\alpha\beta}^\ddagger$ by setting

$$r_{\beta\alpha} = r_0 e^{-\frac{G_{\beta\alpha}^\ddagger - G(\alpha)}{RT}} \quad (10)$$

A short computation then yields

$$G_{\beta\alpha}^\ddagger = -RT \ln \sum_{y \in \beta} \sum_{x \in \alpha} e^{-\frac{E_{xy}^\ddagger}{RT}} \quad (11)$$

as one would expect. This allows us to redraw the barrier tree (which was given in terms of the energies of meta-stable states and their connecting saddle points) in terms of free energies of the corresponding macro-states and their transition states.

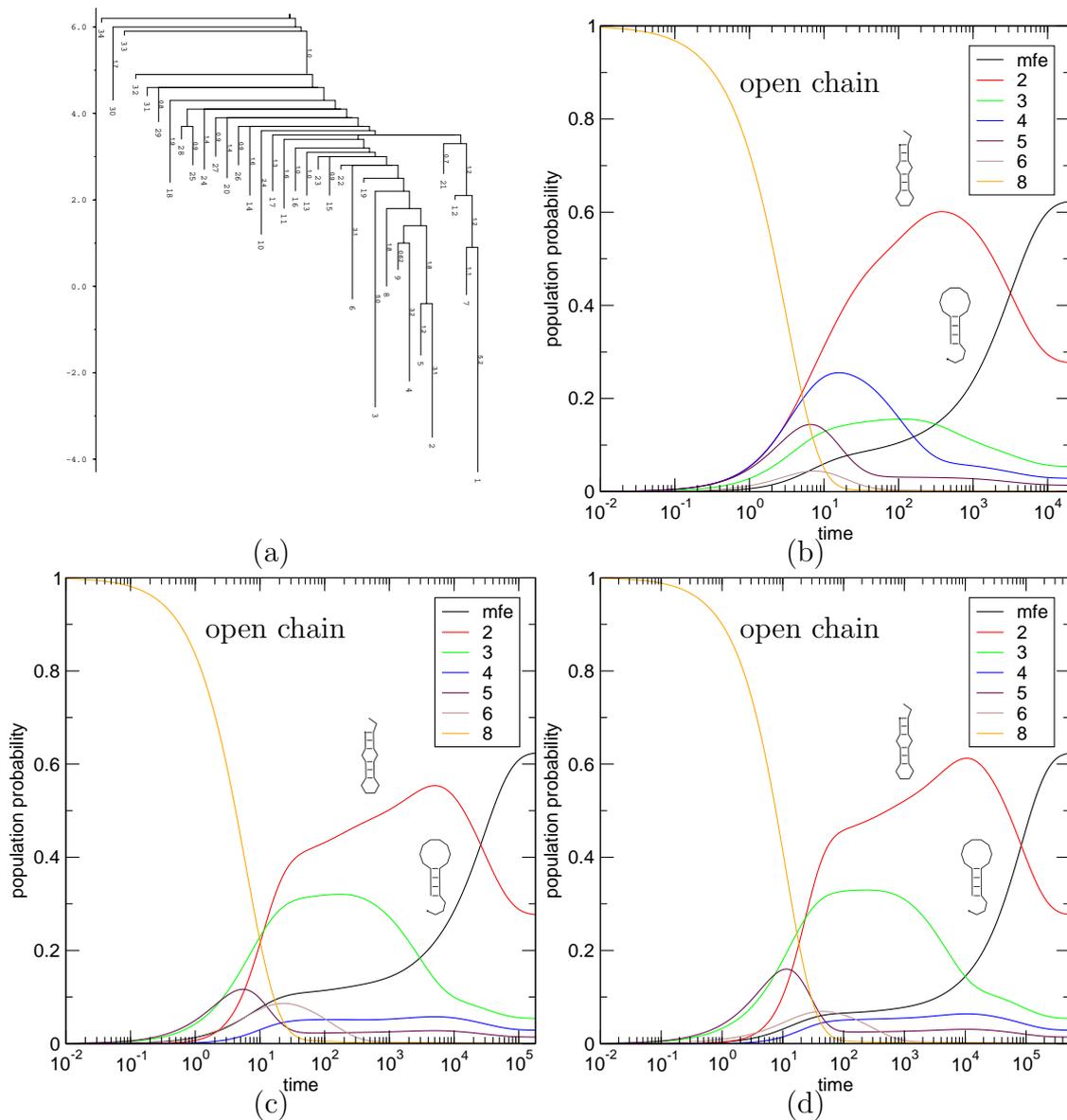


Figure 1. Folding dynamics of xbig (CUGCGGCUUUGGCUCUAGCC). The process was started in the open chain state and run until convergence to the thermodynamic equilibrium distribution. (a) barrier tree, (b) Arrhenius approximation, (c) macro-state process, (d) full process

5. Computational Examples

To demonstrate the quality of the coarse graining above, we present here the folding dynamics for two examples: a short artificial sequence, as well as the well known yeast tRNA^{phe} sequence.

The short sequence, called xbig, has a length of 20 nt and the complete conformation space consists of 3886 secondary structures. For this short sequence it is possible to directly integrate the master equation (2) of the microscopic process and to compare it with the coarse grained dynamics on the space of 34 macro-states corresponding to the

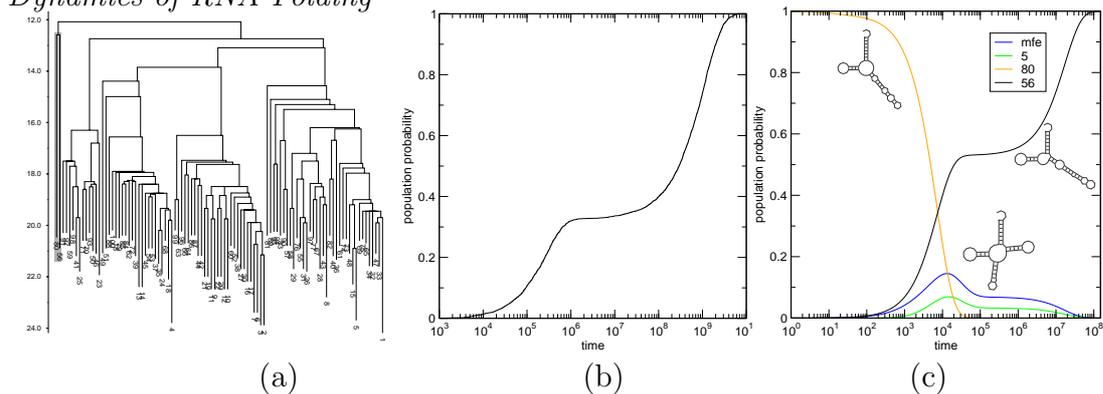


Figure 2. Refolding of a tRNA molecule. The transition from meta-stable state 80 to 56 was chosen because it shows a pronounced plateau indicating two very different pathways: some trajectories cross the highest saddle point and enter the subtree on the right while others refold directly from 80 to 56. (a) barrier tree showing the 100 lowest energy local minima, the subtree containing local minima 56 and 80 is highlighted in red (b) cumulative distribution of first passage times (average of 9000 `kinfold` simulation) (c) occupancies of macro-states computed with an absorbing state attached to basin 56.

local minima of the energy surface. Figure 1 shows that there is excellent agreement between the macro-state approximation (c) and the full process (d). The Arrhenius law gives a qualitatively correct description of the process, although quantitative details are significantly different.

The tRNA sequence, with a length of 76 nt, has some $2.8 \cdot 10^{17}$ possible secondary structures. To recover all saddle points between low-lying local minima we considered the approximately 25 million structures within 15kcal/mol of the ground state, and used `barriers` to compute the 1000 lowest energy local minima as well as the rates between the corresponding macro-states. Only minima with a depth of at least 1kcal/mol were considered in the process.

Obviously, solving the master equation of the full process including the dynamics of all allowed secondary structures is out of the question for such a large conformation space. Instead we compared our coarse grained dynamics to a stochastic sample of trajectories generated by `kinfold`. The program simulates the Markov chain (2) by a rejection-less Monte Carlo algorithm [12]. Further details on the `kinfold` algorithm can be found in [9].

Computing the occupancy of each macro-state from `kinfold` trajectories is very expensive in terms of computer resources, in particular because the time to equilibration becomes too long. Instead we have used `kinfold` to compute first passage times by defining a stop structure in addition to the start point of each trajectory. For the macro-state Markov process we introduce an additional absorbing state Ω , that is accessible only from the macro-state ω that contains the stop structure u with a rate $r_{\Omega\omega} = r_0 e^{-E_u/RT} / Z_\omega$.

The `kinfold` simulations for Fig. 2 required about 3 months of CPU time on an

Intel Pentium 4 running at 2.4 GHz under Linux. In the coarse-grained model, the computational bottleneck concerning CPU and memory resources is the diagonalization of the transition matrix \mathbf{R} , necessary for the computation of $\exp(t\mathbf{R})$. For 1000 states diagonalization takes on the order of 1 minute.

As shown in Fig. 2 simulation and macro-state approximation are in reasonable agreement. The time scale of the macro-state process is shifted somewhat to shorter times and the percentage of trajectories that fold directly is overestimated. This is probably a consequence of the truncation of the energy landscape which leads to incomplete sampling of high energy structures that are more likely to lead outside the 56-80 subtree. For transitions with lower energy barriers the agreement is generally better.

6. Concluding Remarks

We have shown here that a discrete model of secondary structure folding is capable of describing the folding dynamics at macroscopic time-scales that are beyond the reach of methods that operate at atomic resolution. For toy examples one can simply integrate the master equation of the folding dynamics, as has been done also in [46], albeit not using the standard RNA energy model. A controlled approximation to macro-states, defined here as the gradient basins of metastable states, makes the computation of the dynamics feasible for sequences of at least the size of tRNAs (76 nucleotides and approximately 2500 atoms).

The macro-state approximation provides an efficient means of predicting whether a given RNA sequence can act as an RNA switch, and if so, at which time scales.

Often RNA switches are triggered by binding of other molecules. In many cases the interaction partner is another RNA. The folding landscape of a pair of interacting RNAs could be computed explicitly using a modified version of the `RNAsubopt` program, so that at least the case of RNA triggered RNA switches can be modeled entirely within the macro-state approximation. In other cases the energetics of the interaction with the trigger has to be described separately.

Since the `barriers` program used to compute the transition rate matrix \mathbf{R} can deal with arbitrary discrete landscapes, the approach is readily applicable to other problems such as lattice protein folding.

Acknowledgments

This project has been partly funded by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, projects FWF 15893 and by the Austrian Gen-AU bioinformatics integration network sponsored by BM-BWK, and the Bioinformatics Initiative of the DFG, BIZ-6/1-2.

References

- [1] G. T. Barkema and N. Mousseau. Event-based relaxation of continuous disordered systems. *Phys. Rev. Lett.*, 77:4358–4361, 1996.
- [2] T. Baumstark, A. R. Schroder, and D. Riesner. Viroid processing: Switch from cleavage to ligation is driven by a change from a tetraloop to a loop E conformation. *EMBO J.*, 16:599–610, 1997.
- [3] O. M. Becker and M. Karplus. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.*, 106:1495–1517, 1997.
- [4] R. R. Breaker. Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.*, 13:31–39, 2002.
- [5] P. Brion and E. Westhof. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, 26:113–137, 1997.
- [6] B. S. C. Simmerling and A. E. Roitberg. All-atom structure prediction and folding simulations of a stable protein. *JACS*, 124:11258–11259, 2002.
- [7] J. Cupal, I. L. Hofacker, and P. F. Stadler. Dynamic programming algorithm for the density of states of RNA secondary structures. In R. Hofstädt, T. Lengauer, M. Löffler, and D. Schomburg, editors, *Computer Science and Biology 96 (Proceedings of the German Conference on Bioinformatics)*, pages 184–186, Leipzig, Germany, 1996. Universität Leipzig.
- [8] S. A. W. D. Thirumalai, N. Lee and D. K. Klimov. Early events in RNA folding. *Annu. Rev. Phys. Chem.*, 52:751–762, 2001.
- [9] C. Flamm, W. Fontana, I. Hofacker, and P. Schuster. RNA folding kinetics at elementary step resolution. *RNA*, 6:325–338, 2000.
- [10] C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, 216:155–173, 2002.
- [11] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.
- [12] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403, 1976.
- [13] A. P. Gulyaev, F. H. D. vanBatenburg, and C. W. A. Pleij. The computer simulation of RNA folding pathways using a genetic algorithm. *J. Mol. Biol.*, 250:37–51, 1995.
- [14] I. L. Hofacker. The Vienna RNA secondary structure server. *Nucl. Acids Res.*, 31:3429–3431, 2003.
- [15] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125:167–188, 1994.
- [16] I. L. Hofacker, P. Schuster, and P. F. Stadler. Combinatorics of RNA secondary structures. *Discr. Appl. Math.*, 88:207–237, 1998.
- [17] C. Jacob, N. Breton, and P. Daegelen. Stochastic theories of the activated complex and the activated collision: The RNA example. *J. Chem. Phys.*, 107:2903–2912, 1997.
- [18] J. L. Klepeis, H. D. Schafroth, K. M. Westerberg, and C. A. Floudas. Computational methods for protein folding. In R. A. Friesner, editor, *Deterministic Global Optimization and ab Initio Approaches for the Structure Prediction of Polypeptides, Dynamics of Protein Folding, and Protein-Protein Interactions*, volume 120 of *Advances in Chemical Physics*. John Wiley & Sons, New York, 2001.
- [19] T. Klotz and S. Kobe. “Valley structures” in the phase space of a finite 3D Ising spin glass with $\pm i$ interactions. *J. Phys. A: Math. Gen.*, 27:L95–L100, 1994.
- [20] H. M. Martinez. An RNA folding rule. *Nucl. Acid. Res.*, 12:323–335, 1984.
- [21] D. Mathews, J. Sabina, M. Zuker, and H. Turner. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.
- [22] J. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

- [23] A. Mironov and A. Kister. A kinetic approach to the prediction of RNA secondary structures. *J. Biomol. Struct. Dyn.*, 2:953–962, 1985.
- [24] J. H. A. Nagel and C. W. A. Pleij. Self-induced structural switches in RNA. *Biochimie*, 84:913–923, 2002.
- [25] K. Nemoto. Metastable states of the SK spin glass model. *J. Phys. A*, 21:L287–L294, 1988.
- [26] R. Nussinov and I. Tinoco Jr. Sequential folding of a messenger RNA molecule. *J. Mol. Biol.*, 151:519–533, 1981.
- [27] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: The energy landscape perspective. *Ann. Rev. Phys. Chem.*, 48:545–600, 1997.
- [28] S. B. Ozkan, K. A. Dill, and I. Bahar. Computing the transition state populations in simple protein models. *Biopolymers*, 68:35–46, 2003.
- [29] G. A. Papoian and P. G. Wolynes. The physics and bioinformatics of binding and folding – an energy landscape perspective. *Biopolymers*, 68:333–349, 2003.
- [30] A. T. Perrotta and M. D. Been. A toggle duplex in hepatitis delta virus self-cleaving RNA that stabilizes an inactive and a salt-dependent pro-active ribozyme conformation. *J. Mol. Biol.*, 279:361–373, 1998.
- [31] D. Poerschke. Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix. *Biophys. Chem*, 1:381–386, 1974.
- [32] D. Poerschke and M. Eigen. Co-operative non-enzymic base recognition. 3. Kinetics of the helix-coil transition of the oligoribouridylic-oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic pH. *J. Mol. Biol.*, 62:361–381, 1971.
- [33] J. Ponder and D. A. Case. Force fields for protein simulation. *Adv. Prot. Chem.*, 66:27–85, 2003.
- [34] R. A. Poot, N. V. Tsareva, I. V. Boni, and J. van Duin. RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc. Natl. Acad. Sci. USA*, 94(19):10110–10115, 1997.
- [35] R. Rammal, G. Toulouse, and M. A. Virasoro. Ultrametricity for physicists. *Rev. Mod. Phys.*, 58:765–788, 1986.
- [36] M. Schmitz and G. Steger. Description of RNA folding by simulated annealing. *J. Mol. Biol.*, 225:254–266, 1996.
- [37] E. A. Schultes and B. D. P. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452, 2000.
- [38] M. Tacker, W. Fontana, P. F. Stadler, and P. Schuster. Statistics of RNA melting kinetics. *Eur. Biophys. J.*, 23:29–38, 1994.
- [39] A. M. Vertechi and M. A. Virasoro. Energy barriers in SK spin glass models. *J. Phys. France*, 50:2325–2332, 1989.
- [40] B. Voß and R. Giegerich. Prediction of conformational switching in RNA. In H.-W. Mewes, V. Heun, D. Frishman, and S. Kramer, editors, *Proceedings of the German Conference on Bioinformatics 2003*, volume I, pages 173–178, Munich, 2003. belleville Verlag Michael Farin.
- [41] D. J. Wales, J. P. K. Doye, M. A. Miller, P. N. Mortenson, and T. P. Walsh. Energy landscapes: from clusters to biomolecules. *Adv. Chem. Phys.*, 115:3–111, 2000.
- [42] D. J. Wales, M. A. Miller, and T. R. Walsh. Archetypal energy landscapes. *Nature*, 394:758–760, 1998.
- [43] M. S. Waterman. Secondary structure of single-stranded nucleic acids. *Adv. Math. Suppl. Studies*, 1:167 – 212, 1978.
- [44] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49:145–165, 1999.
- [45] H. Zamora, R. Luce, and C. K. Biebricher. Design of artificial short-chained RNA species that are replicated by Q β replicase. *Biochemistry*, 34:1261–1266, 1995.
- [46] W. Zhang and S.-J. Chen. RNA hairpin-folding kinetics. *Proc. Natl. Acad. Sci. USA*, 99:1931–1936, 2002.
- [47] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52, 1989.
- [48] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull. Math. Biol.*,

46:591–621, 1984.