Efficient computation of RNA folding dynamics

2004 J. Phys. A: Math. Gen. 37 4731

(http://iopscience.iop.org/0305-4470/37/17/005)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **37** (2004) 4731–4741 PII: S0305-4470(04)73887-1

Efficient computation of RNA folding dynamics

Michael T Wolfinger[1], W Andreas Svrcek-Seiler[1], Christoph Flamm[1], Ivo L Hofacker[1] and Peter F Stadler[2]

[1] Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, Währingerstraße 17, A-1090 Wien, Austria
[2] Bioinformatik, Institut für Informatik, Universität Leipzig, D-04103 Leipzig, Germany

E-mail: ivo@tbi.univie.ac.at

Received 22 December 2003, in final form 8 March 2004
Published 14 April 2004
Online at stacks.iop.org/JPhysA/37/4731 (DOI: 10.1088/0305-4470/37/17/005)

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 behaviour 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 behaviour of RNA switches.

PACS numbers: 87.14.Gg, 87.15.He, 87.15.Aa, 87.15.Cc

1. Introduction

A 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 of at least the size of tRNAs can be computed for arbitrarily long timescales 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 also used with some success 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 for 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 metastable multi-component structure and a rod-like conformation, constituting the native state, separated by a huge energy barrier. While the metastable 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 metastable 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 in 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 favour 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 of 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 acts as an antidote by binding to the activated *hok* mRNA, leading to a 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 refer to [4, 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 timescales. The simulated annealing approach to secondary structure formation used in [36] is based on the same idea. Other methods for simulating 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 $E(x) \leqslant E(y)$ for all its neighbours, $(x, y) \in \mathcal{M}$. A gradient walk is defined as follows: starting from $x \in X$ we move to its neighbour y with minimal energy if $E(y) < E(x)$. If the minimum energy neighbour 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 $\gamma : X \to X$ and each $x \in X$ is mapped onto 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 onto 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

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

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

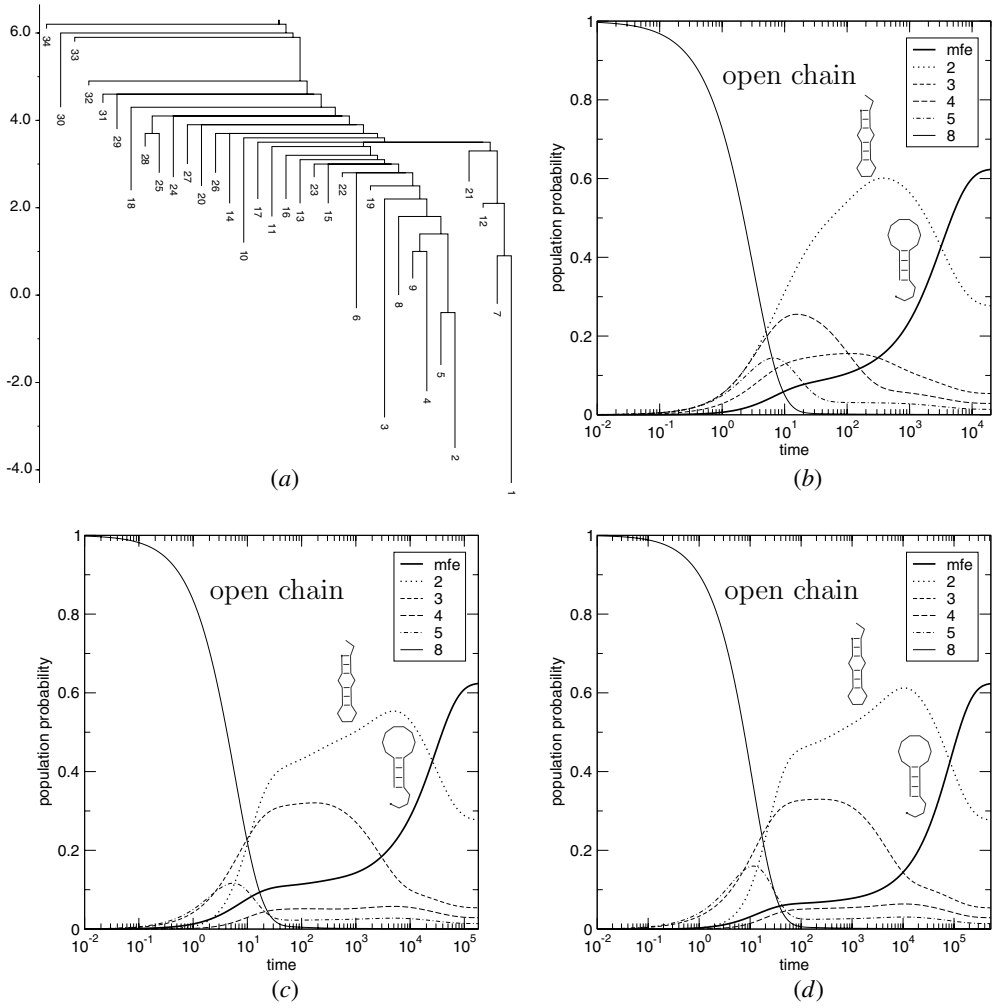

Figure 1. Folding dynamics of xbix (CUGCGGCUUUGGCUCUAGCC). The process was started in the open chain state and run until convergence to the thermodynamic equilibrium distribution. (*a*) Barrier tree with local minima numbered by their energy, (*b*) Arrhenius approximation, (*c*) macrostate process, (*d*) full process. The label 'mfe' marks the minimum free energy structure (local min. 1).

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 figure 1(*a*)). Since saddle points separate local optima, each valley $\mathcal{V}(s)$ contains (in the non-degenerate case at least two) local minima $z_1, \ldots, z_k$. 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 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 equation (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 an earlier bold study of the energy barriers of the SK model [25], see also [19]. Instead, the program package, `barriers`,[3] 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 for 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 favoured over long-distance structural elements, because of the spatial proximity of the opposing base-pair partners [5, 26].

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 modelling conformational changes in terms of elementary steps of opening and closing of base pairs. In this approach the transition rate r_{xy} from the secondary structure y to the secondary structure x is nonzero only if $(y, x) \in \mathcal{M}$, i.e. if x and y are neighbours 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

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = \sum_{y \in X} r_{xy}\, p_y(t) \qquad \text{with} \quad r_{xx} = -\sum_{y \neq x} r_{yx}. \tag{2}$$

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 $\vec{p}(t) = \exp(t\mathbf{R})\,\vec{p}(0)$.

[3] The software is available from http://www.tbi.univie.ac.at/∼ ivo/RNA/Barriers.

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

$$r_{yx} = r_0 \exp\left(-\frac{E^{\neq}_{yx} - E(x)}{RT}\right) \qquad \text{for} \quad x \neq y \tag{3}$$

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

$$E^{\neq}_{yx} = \max\{E(x), E(y)\} \tag{4}$$

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. Macrostates 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 $\Pi = \{\alpha, \beta, \ldots\}$ be a partition of the state space X. The classes of such a partition are *macrostates*. 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

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

and the corresponding free energy

$$G(\alpha) = -RT \ln Z_\alpha. \tag{6}$$

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

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

where $\mathrm{Prob}[x|\alpha]$ is the probability of occupying state $x \in \alpha$ given that we know the process is in macrostate α. The kinetics of the molecule in terms of its macrostates is given by the master equation

$$\frac{\mathrm{d}p_\alpha}{\mathrm{d}t} = \sum_{\beta\in\mathbf{\Pi}} r_{\alpha\beta}\, p_\beta(t) \tag{8}$$

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 $\mathrm{Prob}[x|\alpha] = \mathrm{e}^{-E(x)/RT}/Z_\alpha$ and hence

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

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 macrostate it belongs and (b) the double sum in equation (9) needs to be evaluated only for neighbouring 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 neighbours.

Condition (a) is easily satisfied for each of the gradient basins: in each step of the `barriers` algorithm all neighbours 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 neighbour. Initially, each local optimum forms the nucleus of new gradient basin, hence the macrostate to which x belongs can be determined in $\mathcal{O}(\delta)$ operations, where δ is the maximum number of neighbours of a secondary structure.

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

$$r_{\beta\alpha} = r_0 \exp\left(-\frac{G^{\neq}_{\beta\alpha} - G(\alpha)}{RT}\right). \tag{10}$$

A short computation then yields

$$G^{\neq}_{\beta\alpha} = -RT \ln \sum_{y\in\beta} \sum_{x\in\alpha} \exp\left(-\frac{E^{\neq}_{xy}}{RT}\right) \tag{11}$$

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

5. Computational examples

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

The short sequence, called xbix, 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 macrostates corresponding to the local minima of the energy surface. Figure 1 shows that there is excellent agreement between the macrostate 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 2.8×10^{17} possible secondary structures. To recover all saddle points between low-lying local minima, we considered approximately 25 million structures within 15 kcal mol^{-1} of the ground state, and used `barriers` to compute the 1000 lowest energy local minima as well as the rates between the corresponding macrostates. Only minima with a depth of at least 1 kcal mol^{-1} 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 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].

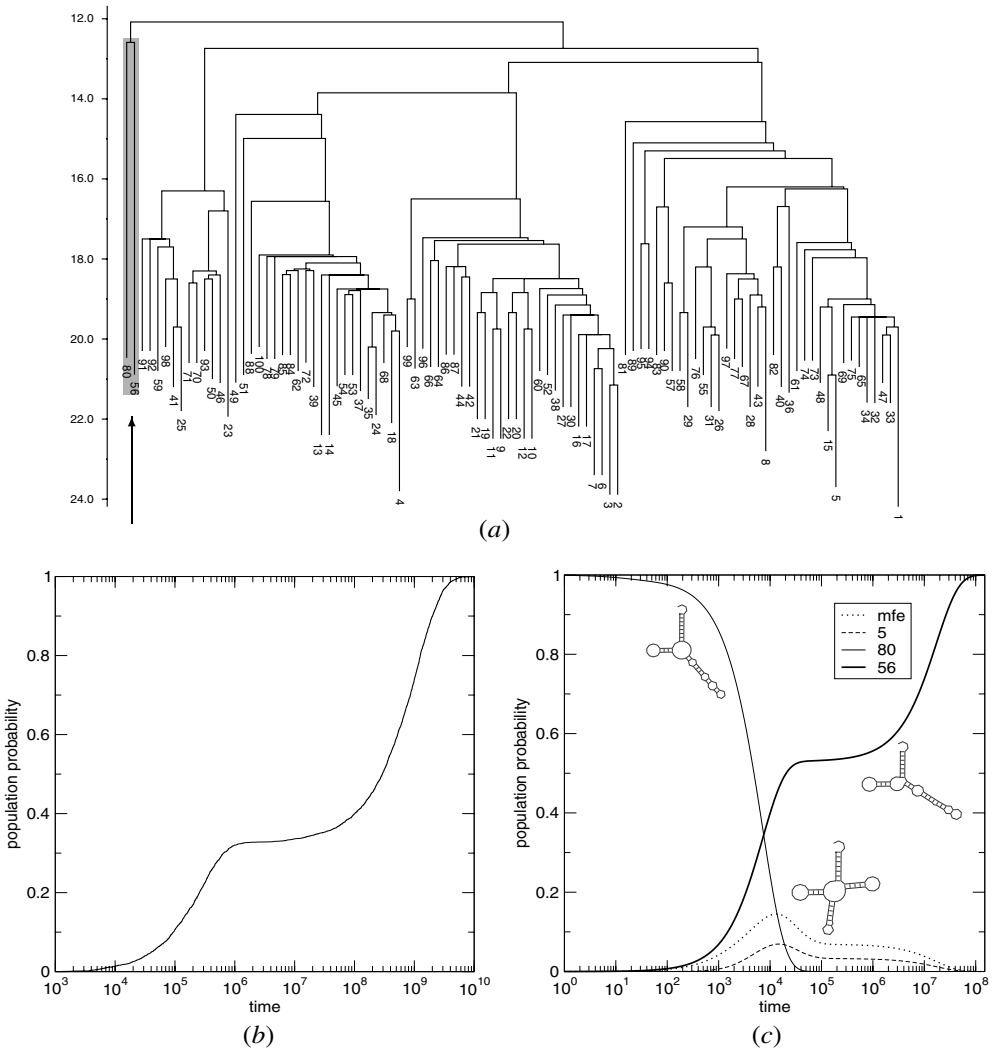

Figure 2. Refolding of a tRNA molecule. The transition from metastable state 80 to 56 was chosen because it shows a pronounced plateau indicating two very different pathways: some trajectories cross the highest saddle point and enter the subtree on the right while others refold directly from 80 to 56. (*a*) Barrier tree showing the 100 lowest energy local minima, local minima 56 and 80 are the two left-most states (highlighted in grey and marked by an arrow); (*b*) cumulative distribution of first passage times (average of 9000 `kinfold` simulations) and (*c*) occupancies of macrostates 1, 5, 56 and 80, computed with an absorbing state attached to basin 56.

Computing the occupancy of each macrostate 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 macrostate Markov process, we introduce an additional absorbing state, Ω, that is accessible only from the macrostate ω, which contains the stop structure u with a rate $r_{\Omega\omega} = r_0 \exp(-E_u/RT)/Z_\omega$.

The `kinfold` simulations for figure 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 the 1000 states used in the example, diagonalization takes on the order of 1 min.

As shown in figure 2 simulation and macrostate approximation are in reasonable agreement. The timescale of the macrostate 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 an 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 timescales 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 also been done in [46], albeit not using the standard RNA energy model. A controlled approximation to macrostates, 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 macrostate approximation provides an efficient means of predicting whether a given RNA sequence can act as an RNA switch, and if so, at which timescales.

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 modelled entirely within the macrostate 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] Barkema G T and Mousseau N 1996 Event-based relaxation of continuous disordered systems *Phys. Rev. Lett.* **77** 4358–61

[2] Baumstark T, Schroder A R and Riesner D 1997 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

[3] Becker O M and Karplus M 1997 The topology of multidimensional potential energy surfaces: theory and application to peptide structure and kinetics *J. Chem. Phys.* **106** 1495–517

[4] Breaker R R 2002 Engineered allosteric ribozymes as biosensor components *Curr. Opin. Biotechnol.* **13** 31–9

[5] Brion P and Westhof E 1997 Hierarchy and dynamics of RNA folding *Annu. Rev. Biophys. Biomol. Struct.* **26** 113–37

[6] Simmerling B S C and Roitberg A E 2002 All-atom structure prediction and folding simulations of a stable protein *J. Am. Chem. Soc.* **124** 11258–9
[7] Cupal J, Hofacker I L and Stadler P F 1996 Dynamic programming algorithm for the density of states of RNA secondary structures *Computer Science and Biology 96: Proc. German Conf. on Bioinformatics* ed R Hofstädt, T Lengauer, M Löffler and D Schomburg *(Leipzig, Germany, 1996)* (Leipzig: Universität Leipzig) pp 184–6
[8] Thirumalai S A W D, Lee N and Klimov D K 2001 Early events in RNA folding *Annu. Rev. Phys. Chem.* **52** 751–62
[9] Flamm C, Fontana W, Hofacker I and Schuster P 2000 RNA folding kinetics at elementary step resolution *RNA* **6** 325–38
[10] Flamm C, Hofacker I L, Stadler P F and Wolfinger M T 2002 Barrier trees of degenerate landscapes *Z. Phys. Chem.* **216** 155–73
[11] Frauenfelder H, Sligar S G and Wolynes P G 1991 The energy landscapes and motions of proteins *Science* **254** 1598–603
[12] Gillespie D T 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions *J. Comput. Phys.* **22** 403
[13] Gultyaev A P, vanBatenburg F H D and Pleij C W A 1995 The computer simulation of RNA folding pathways using a genetic algorithm *J. Mol. Biol.* **250** 37–51
[14] Hofacker I L 2003 The Vienna RNA secondary structure server *Nucleic. Acids Res.* **31** 3429–31
[15] Hofacker I L, Fontana W, Stadler P F, Bonhoeffer L S, Tacker M and Schuster P 1994 Fast folding and comparison of RNA secondary structures *Monatsh. Chem.* **125** 167–88
[16] Hofacker I L, Schuster P and Stadler P F 1998 Combinatorics of RNA secondary structures *Discrete. Appl. Math.* **88** 207–37
[17] Jacob C, Breton N and Daegelen P 1997 Stochastic theories of the activated complex and the activated collision: the RNA example *J. Chem. Phys.* **107** 2903–12
[18] Klepeis J L, Schafroth H D, Westerberg K M and Floudas C A 2001 Computational methods for protein folding *Deterministic Global Optimization and ab initio Approaches for the Structure Prediction of Polypeptides, Dynamics of Protein Folding, and Protein-Protein Interactions (Advances in Chemical Physics* vol 120*)* ed R A Friesner (New York: Wiley)
[19] Klotz T and Kobe S 1994 '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
[20] Martinez H M 1984 An RNA folding rule *Nucleic Acids Res.* **12** 323–35
[21] Mathews D, Sabina J, Zuker M and Turner H 1999 Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure *J. Mol. Biol.* **288** 911–40
[22] McCaskill J 1990 The equilibrium partition function and base pair binding probabilities for RNA secondary structure *Biopolymers* **29** 1105–19
[23] Mironov A and Kister A 1985 A kinetic approach to the prediction of RNA secondary structures *J. Biomol. Struct. Dyn.* **2** 953–62
[24] Nagel J H A and Pleij C W A 2002 Self-induced structural switches in RNA *Biochimie* **84** 913–23
[25] Nemoto K 1988 Metastable states of the SK spin glass model *J. Phys. A: Math. Gen.* **21** L287–L294
[26] Nussinov R and Tinoco I Jr 1981 Sequential folding of a messenger RNA molecule *J. Mol. Biol* **151** 519–33
[27] Onuchic J N, Luthey-Schulten Z and Wolynes P G 1997 Theory of protein folding: the energy landscape perspective *Annu. Rev. Phys. Chem.* **48** 545–600
[28] Ozkan S B, Dill K A and Bahar I 2003 Computing the transition state populations in simple protein models *Biopolymers* **68** 35–46
[29] Papoian G A and Wolynes P G 2003 The physics and bioinformatics of binding and folding—an energy landscape perspective *Biopolymers* **68** 333–49
[30] Perrotta A T and Been M D 1998 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** 361–73
[31] Poerschke D 1974 Thermodynamic and kinetic parameters of an oligonucleotide hairpin helix *Biophys. Chem.* **1** 381–6
[32] Poerschke D and Eigen M 1971 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** 361–81
[33] Ponder J and Case D A 2003 Force fields for protein simulation *Adv. Protein Chem.* **66** 27–85
[34] Poot R A, Tsareva N V, Boni I V and van Duin J 1997 RNA folding kinetics regulates translation of phage MS2 maturation gene *Proc. Natl Acad. Sci. USA* **94** 10110–5
[35] Rammal R, Toulouse G and Virasoro M A 1986 Ultrametricity for physicists *Rev. Mod. Phys.* **58** 765–88

[36] Schmitz M and Steger G 1996 Description of RNA folding by simulated annealing *J. Mol. Biol.* **225** 254–66
[37] Schultes E A and Bartel D P 2000 One sequence, two ribozymes: implications for the emergence of new ribozyme folds *Science* **289** 448–52
[38] Tacker M, Fontana W, Stadler P F and Schuster P 1994 Statistics of RNA melting kinetics *Eur. Biophys. J.* **23** 29–38
[39] Vertechi A M and Virasoro M A 1989 Energy barriers in SK spin glass models *J. Physique* **50** 2325–32
[40] Voß B and Giegerich R 2003 Prediction of conformational switching in RNA *Proc. German Conf. on Bioinformatics 2003* vol I 173–178, ed H-W Mewes, V Heun, D Frishman and S Kramer (Munich: Belleville Michael Farin)
[41] Wales D J, Doye J P K, Miller M A, Mortenson P N and Walsh T P 2000 Energy landscapes: from clusters to biomolecules *Adv. Chem. Phys.* **115** 3–111
[42] Wales D J, Miller M A and Walsh T R 1998 Archetypal energy landscapes *Nature* **394** 758–60
[43] Waterman M S 1978 Secondary structure of single-stranded nucleic acids *Adv. Math. Suppl. Studies* **1** 167–212
[44] Wuchty S, Fontana W, Hofacker I L and Schuster P 1999 Complete suboptimal folding of RNA and the stability of secondary structures *Biopolymers* **49** 145–65
[45] Zamora H, Luce R and Biebricher C K 1995 Design of artificial short-chained RNA species that are replicated by Qβ replicase *Biochemistry* **34** 1261–6
[46] Zhang W and Chen S-J 2002 RNA hairpin-folding kinetics *Proc. Natl Acad. Sci. USA* **99** 1931–6
[47] Zuker M 1989 On finding all suboptimal foldings of an RNA molecule *Science* **244** 48–52
[48] Zuker M and Sankoff D 1984 RNA secondary structures and their prediction *Bull. Math. Biol.* **46** 591–621