Chapter 25
Finding Attractors on a Folding Energy Landscape

Wilfred Ndifon
Princeton University, USA & Weizmann Institute of Science, Israel

Jonathan Dushoff
McMaster University, Canada

ABSTRACT

RNA sequences fold into their native conformations by means of an adaptive search of their folding energy landscapes. The energy landscape may contain one or more suboptimal attractor conformations, making it possible for an RNA sequence to become trapped in a suboptimal attractor during the folding process. How the probability that an RNA sequence will find a given attractor before it finds another one depends on the relative positions of those attractors on the energy landscape is not well understood. Similarly, there is an inadequate understanding of the mechanisms that underlie differences in the amount of time an RNA sequence spends in a particular state. Elucidation of those mechanisms would contribute to the understanding of constraints operating on RNA folding. This chapter explores the kinetics of RNA folding using theoretical models and experimental data. Discrepancies between experimental predictions and expectations based on prevailing assumptions about the determinants of RNA folding kinetics are highlighted. An analogy between kinetic accessibility and evolutionary accessibility is also discussed.

INTRODUCTION

Biopolymers such as proteins and RNA play significant functional roles in living organisms; they are crucial for, among other things, locomotion, protection against disease, regulation of gene expression, and catalysis of biochemical reactions (Alberts *et al*., 2002; Yen *et al*., 2004; Serganov *et al*., 2006). This functionality is often mediated by specific native conformations. The attainment of such native conformations involves a search of an astronomically large conformation space, via a series of elementary structural rearrangements. At

DOI: 10.4018/978-1-60960-491-2.ch025

the molecular level, these elementary structural rearrangements include the formation/dissociation of hydrogen bonds between ribonucleotides and the formation/dissociation of hydrogen, van der Waals, and disulfide bonds between amino acid residues. If the search for the native conformation were a simple random process, it would take a biopolymer a biologically unrealistic length of time to find its native conformation following its biosynthesis. For example, assuming that during folding an RNA stays at each of its possible non-native secondary structures no longer than the duration of an atomic oscillation, a simple calculation shows that it would take $\sim 4.5 \times 10^6$ years for the phenylalanil-tRNA from the yeast *Saccharomyces cereviseae* to complete a random search for its native conformation. This discrepancy between the time scale required for a random search of the conformation space and the time scale required for biological functionality was originally pointed out by Levinthal (1969) and is known as Levinthal's paradox.

There is, however, evidence that the search for the native conformation is not random but it proceeds along an energy gradient induced by differences in the thermal stabilities of the possible biopolymer conformations (Onuchic *et al*., 2000, Wolynes, 2005). A folding biopolymer "moves" on an "energy landscape", from regions of relatively high free energy to regions of lower free energy. A number of models applying this energy landscape perspective to analyze the RNA folding process have been published. Molecular dynamics models (Pan & Mackerell, 2003; Sorin *et al*., 2004) are mostly suitable only for the analysis of very short RNA sequences due to the very large number of computations involved. Analytical models (Zhang & Chen, 2002, 2006) are similarly limited in the scope of their application because they require the enumeration of an astronomical number of possible conformations of the RNA sequence under consideration. Monte Carlo models (Flamm *et al*., 2000; Isambert & Siggia, 2000; Xayaphoumine *et al*., 2003; Ndifon, 2005) and models based on genetic algorithms (Gultyaev *et al*., 1995; Shapiro *et al*., 2001) partially circumvent these computational limitations and they are consequently applicable to RNA sequences of varying lengths. Important drawbacks of the genetic algorithms include the facts that (1) they do not allow the inference of physically relevant RNA folding times, and (2) they employ certain parameters (e.g., mutation and crossover rates) that do not relate to any physical aspects of RNA folding dynamics and whose values are determined by means of trial and error (Higgs, 2000).

There are important open questions concerning the shape of the energy landscapes of natural RNAs and how this shape determines the kinetics of RNA folding. To motivate the particular questions addressed in this chapter, it is useful to recall that some folding paths of certain RNA sequences (e.g., the SV11 RNA sequence) lead to suboptimal attractor conformations that, once realized, take a very long time to unravel. In some cases, the suboptimal attractors are functional within the time window required for their unraveling (e.g., see Gultyaev *et al*., 1995). However, in many cases the suboptimal attractors simply extend the waiting time until an RNA sequence can attain its optimal conformation and thereby become functional. This chapter investigates the above aspects of the kinetics of RNA folding using both theoretical models and experimental data.

RNA FOLDING MODEL

In this section, the model of RNA folding kinetics used in this work is described in detail. First, a working definition of RNA sequences and their conformations is given. Methods for calculating the free energies of RNA conformations conditioned on the underlying RNA sequence are also described. The model of RNA folding kinetics is then presented.