Landscapes and Energy Barriers

Michael Wolfinger, Peter F. Stadler, Ivo L. Hofacker, Christoph Flamm

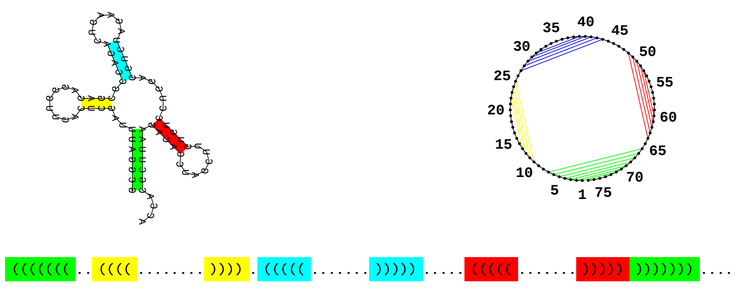
Institute for Theoretical Chemistry and Molecular Structural Biology, University Vienna

Tel: +43 1 4277 52731 Fax: +43 1 4277 52793 Email: {mtw,studla,ivo,xtof}@tbi.univie.ac.at Web: http://www.tbi.univie.ac.at/

A tool for a thorough investigation of RNA energy landscapes is presented. The topological details of these landscapes are represented by so called 'barrier trees', which give an efficient impression of the landscape and its overall shape and characteristics.

RNA Secondary Structure

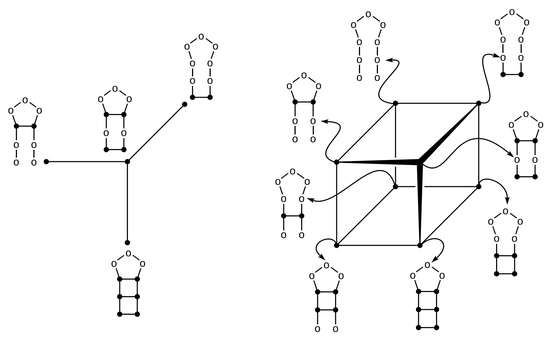
An interesting aspect concerning biomolecules is structure prediction. As RNA folding is thought to be of hierarchical nature [2], secondary structures can be seen as a coarse grained approach to the three dimensional structures.



RNA secondary structure: (left) conventional representation. (right) circle representation. (below) bracket-dot representation.

RNA secondary structure is defined as a pattern of base pairs, which is determined by hydrogen bonds between the four bases Adenine (A), Guanine (G), Cytosine (C) and Uracil (U). Efficient algorithms for calculation and evaluation of RNA secondary structures have been suggested [6, 4]. An important contribution to the understanding of the behavior of RNA molecules was given in [5], where a tool for calculation of all suboptimally folded RNA structures within a certain energy range above the ground state was introduced.

The energy landscape of a RNA molecule is a complex surface of the free energy versus the conformational degrees of freedom. Here, the allowed conformations are the (suboptimal) secondary structures which are compatible with a particular sequence. The figure below illustrates the 'conformation space $\mathcal{C}$' of a short RNA molecule.



One move neighborhood of the conformation space (l.h.s.) and its embedding in the graph representing the conformation space (r.h.s) for a small RNA molecule which can exhibit 3 base pairs.

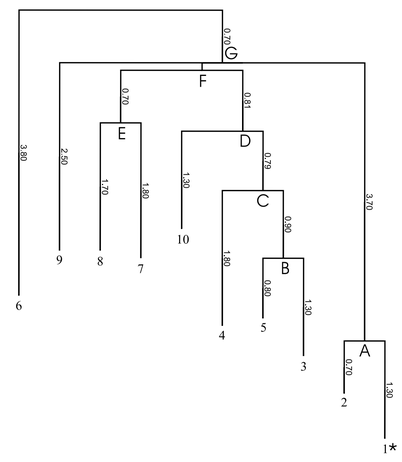
As $\mathcal{C}$ is a multidimensional space, it is not clear a priori how to move in such a complex space. It is therefore necessary to define certain rules, the so called 'move set' (a collection of operations, which, applied to an element of $\mathcal{C}$, transforms this element into another element of $\mathcal{C}$.

Energy Barriers

With both features at hand (all suboptimal structures and a metric, the move set, a more detailed investigation of the energy landscape of RNA is possible. It is therefore necessary make some definitions:

- A structure is a local minimum if its energy is lower than the energy of all neighboring structures
- A structure is called local maximum if its energy is higher than the energies of all legal neighboring structures
- A structure is a saddle point if there are at least two local minima that can be reached by a downhill walk starting with this structure

Evidently, the saddle point with lowest energy separating two local minima is of particular importance. They can be found by applying the flooding algorithm presented below. The outcome of the procedure is a barrier tree as shown in this figure.

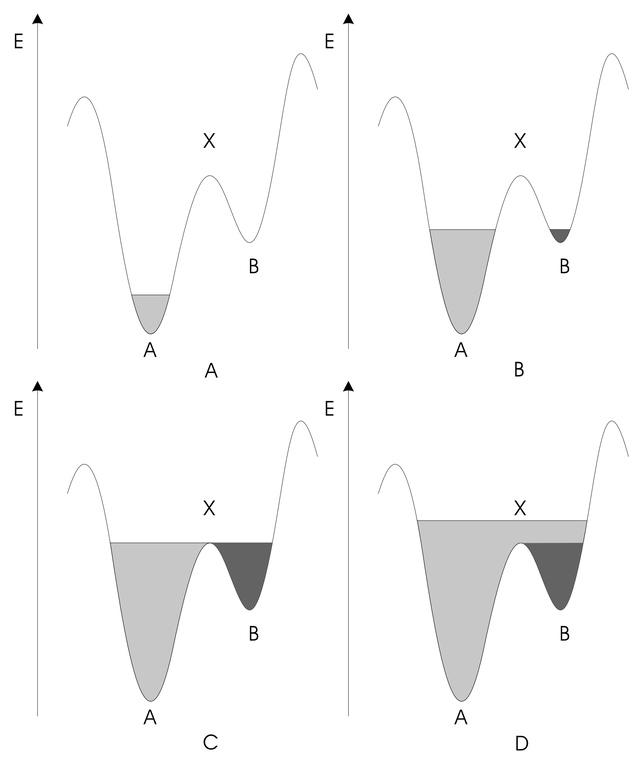


Barrier tree of a random RNA sequence with length 42: Leaves 1-10 denote the 10 lowest local minima of the energy landscape, the mfe structure 1 is marked with an asterisk. Saddle points are labeled A to G.The Energy barrier of 3 is $B(3) = E(B) - E(3)$.

Leaves correspond to the valleys of the landscape, while saddle points are displayed by internal nodes. Saddle points can be read off easily from these barrier trees.

The flooding algorithm

The flooding algorithm can be explained schematic with the following figure.

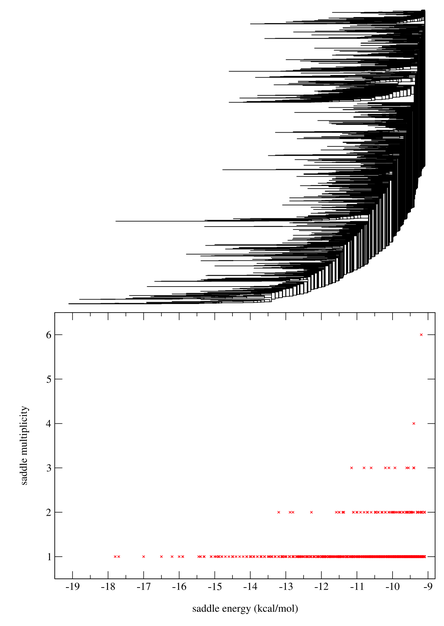


The flooding algorithm: Gedanken-experiment where water rises in a landscape from bottom to top..

Imagine a landscape with only two valleys A and B (where A is energetically lower than B) and a saddle point X separating those local minima. Water rises from bottom to top. In the first step (A), only the deeper valley will be slightly filled with water. For the algorithm, this means that all structures below or exactly at the surface of the water belong to valley A. (All other structures are not accessible by now as we go through an energetically sorted list of secondary structures in ascending order.) In the second step (B), not only the deeper valley A is filled with water, but also valley B (at least to a small amount). From now on there are two possibilities for a structure to belong to: Depending on which valley contains structures that are neighbors of the actual one, a structure can belong to valley A or B. Imagine the water rises further. Step (C) displays a different situation: Saddle point X has been found, which means there exists a structure which has legal neighbors in valley A *and* in valley B. At this point, the two valleys coincide and B is merged with its 'father A, which means that all structures from B can be accessed as if they would belong to A from now on. However, the algorithm does not stop here. As illustrated in (D), the water rises further and only valley A is still accessible. The end of the algorithm has been reaches as soon as (i) all secondary structures have been processed or (ii) a predefined amount of local minima has been found.

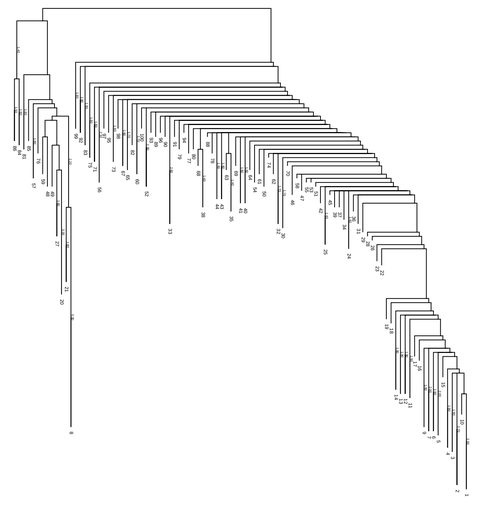
Applications

The figure below shows the frequency of degenerate saddles from a tRNA^{phe}.



Saddle point energy versus saddle point multiplicity from a tRNA^{phe}

Another useful application is shown below: The barrier tree of a bi-stable RNA secondary structure, which can fold into two or more thermodynamically stable structures separaded by a large energy barrier. Recently, artificial RNA switches have been designed.



Barrier tree of a RNA switch showing the 100 deepest local minima of the energy landscape: The deep local minimum 8 (-6.7 kcal/mol) on the left hand side is separated from the right subtree via an energy barrier of 10.10 kcal/mol. The energy of the mfe structure is -8.20 kcal/mol. The open chain conformation is represented by local minimum 81 in the very left part of the tree.

Besides the shown applications, barrier trees have been considered recently for various models of disordered systems, including spin glasses and combinatorial optimization problems [1, 3].

References

[1] O. Bastert, D. Rockmore, P. Stadler, and G. Tinhofer. Landscapes on spaces of trees. Appl. Math. Comput. 2001, in press.

[2] P. Brion and E. Westhof. Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, 26:113–137, 1997.

[3] F. Ferreira, J. Fontanari, and P. Stadler. Landscape statistics of the low autocorrelated binary string proble m. *J. Phys. A: Math. Gen*, 33:8635–8647, 2000.

[4] 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.

[5] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structure. *Biopolymers*, 49:145–165, 1998.

[6] M. Zuker and D. Sankoff. RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46:591–621, 1984.