Kinwalker - Kinetic Backtracking of RNA folding

Michael Geis
Caro Turner,Christoph Flamm,Peter Stadler,Martin Middendorf,Michael Wolfinger

Bioinformatics Group, Dept. of Computer Science
University of Leipzig

Embio Meeting Leipzig, May 18th 2007

1 Introduction

2 Algorithm
- Select set of states
- Find starting structure
- Extending the Front
- Saddle Height

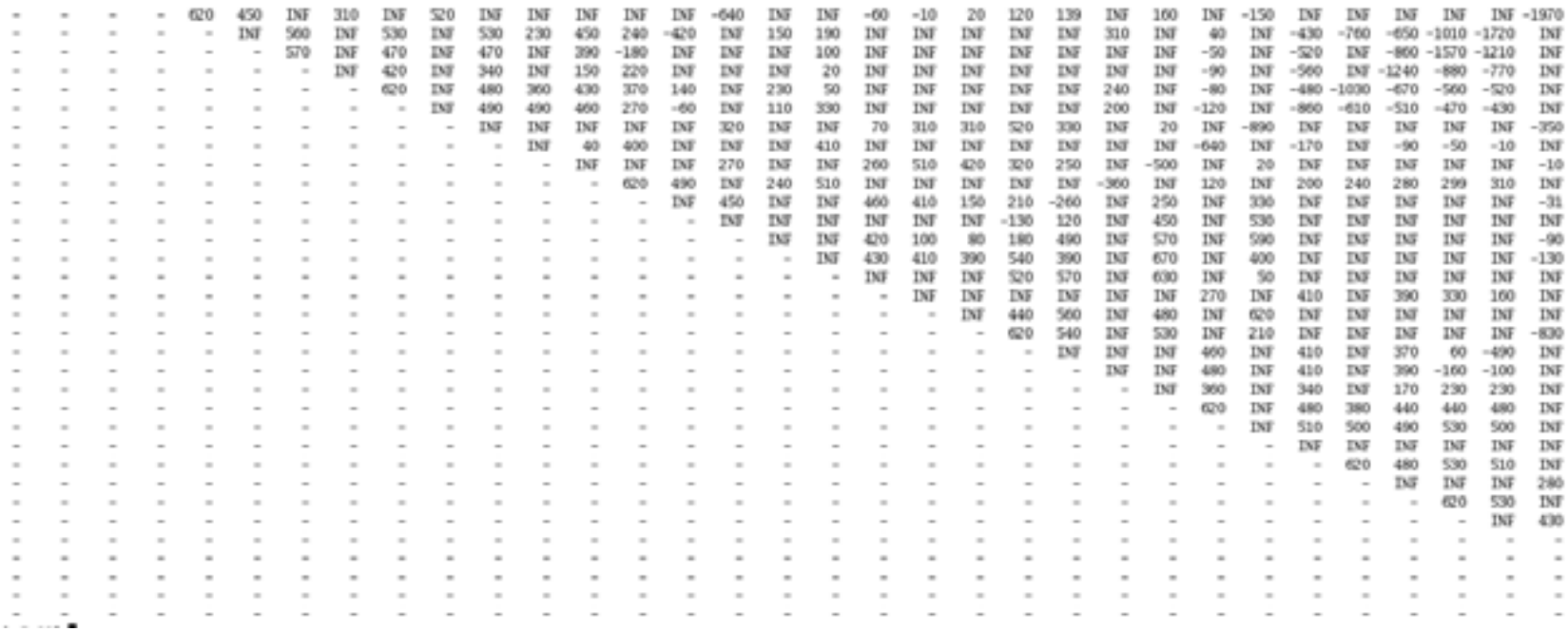
Essential components in simulating RNA folding:

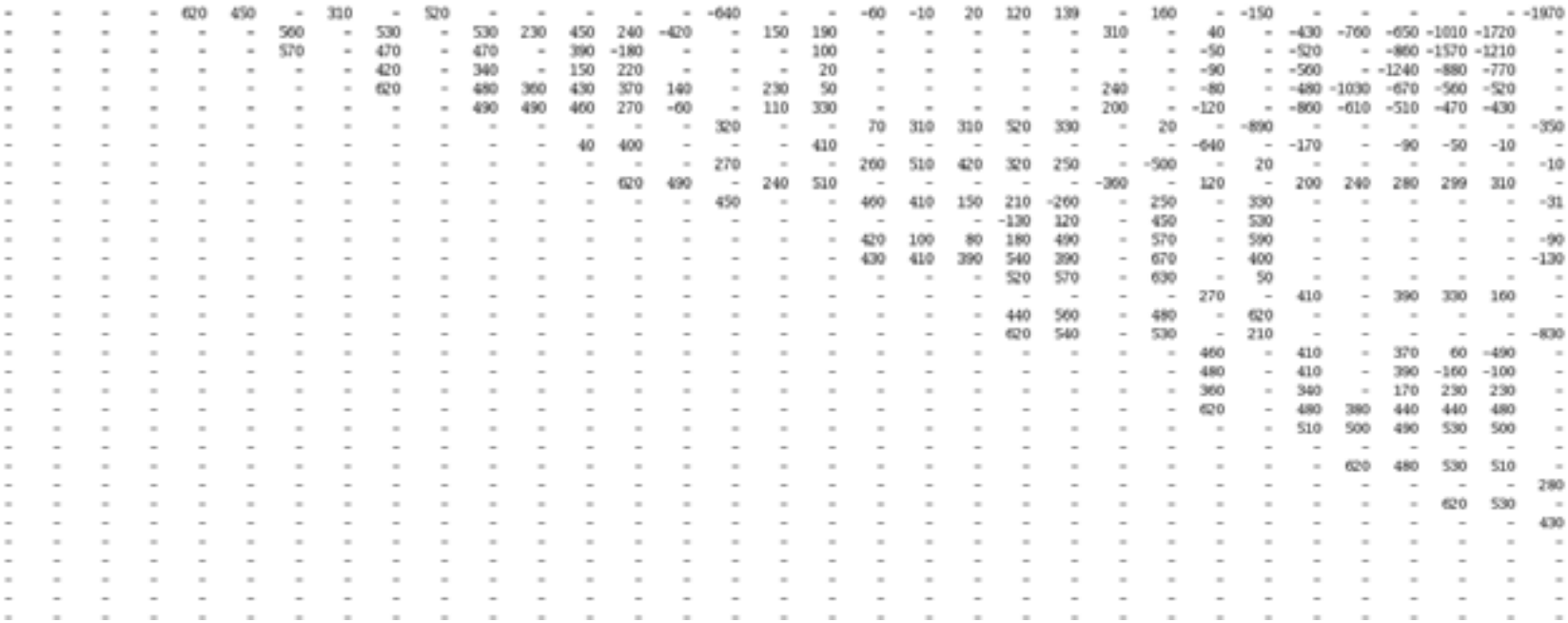
Select a set of states **Model state transitions**

Main issue: How to calculate energy barriers between states.

Calculate the suboptimal structures

For $1 <= i <= j <= n$ consider as states the lowest energy structures on a subsequence (i,j) given that bases i and j pair (extremes).

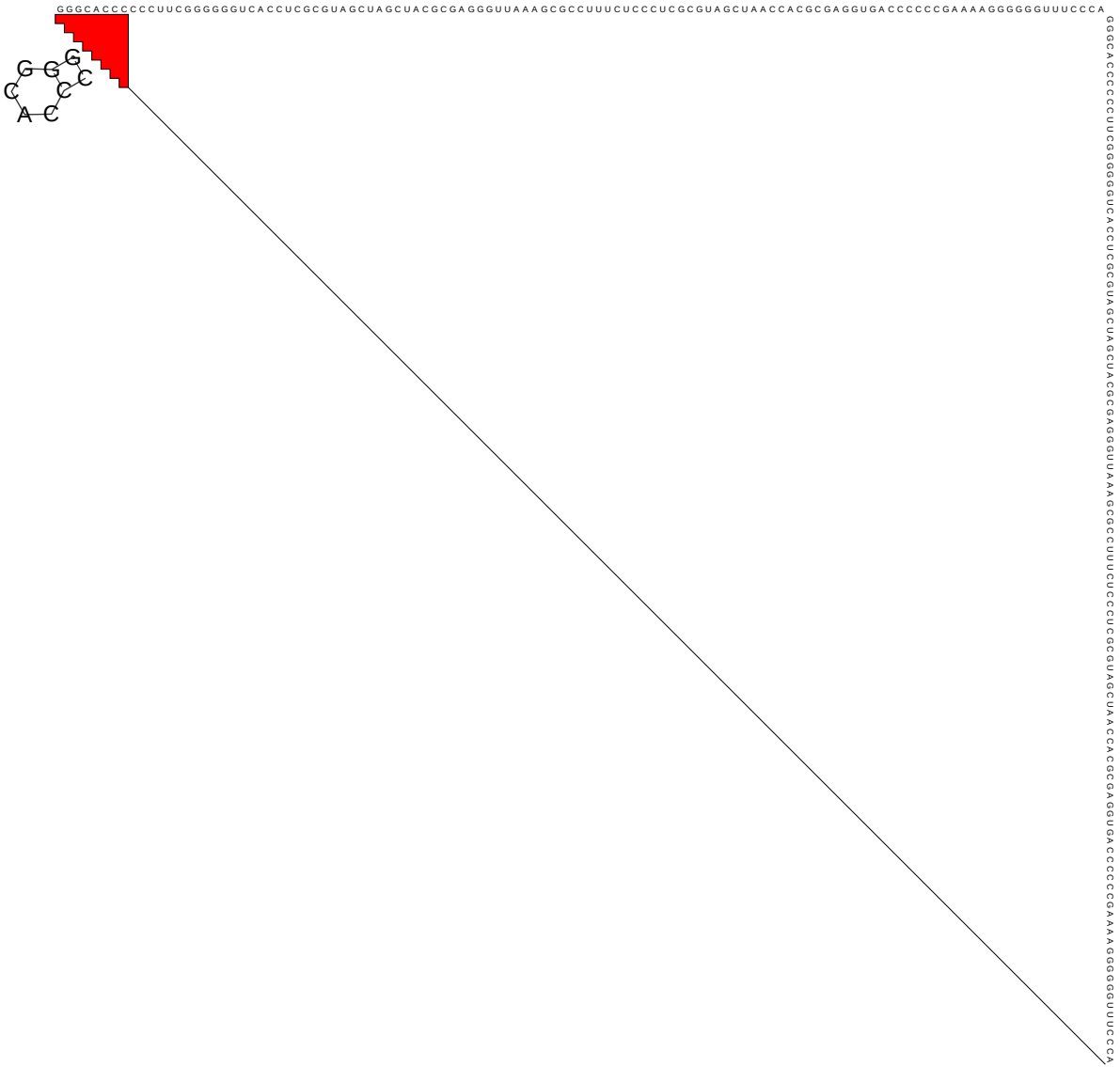




Determine the initial energy front

1. Rank extremes by
   i) diagonal
   ii) distance from edge of matrix
   iii) 5' before 3'
2. Find the diagonal closest to the main diagonal that is not empty
3. Add substructures in that diagonal to the front if they are transcribed and there is no conflict

dot.ps


GGGCACCCCCCUUCGGGGGGUCACCUCGCGUAGCUAGCUACGCGAGGGUUAAAGCGCCUUUCUCCCUCGCGUAGCUAACCACGCGAGGUGACCCCCCGAAAAGGGGGGUUUCCCA

Repeat either of the following (depending on first-passage times) until the mfE structure is reached:

1. Fold into a secondary structure of lower energy
2. Transcribe another base

The first-passage time depends on the energy barrier that has to be crossed to fold into the new state. Whether 1. or 2. is done, therefore, depends on the energy barrier.

Conflict

A substructure (i,j) determines how the bases in that subsequence pair. Once (i,j) is part of the front, and a pair (k,l) overlapping with (i,j) is added, they may demand demand base pairs in the overlapping region.

Conflict Resolution

1. Add only those basepairs of a conflicting substructure that don't conflict with the current structure.

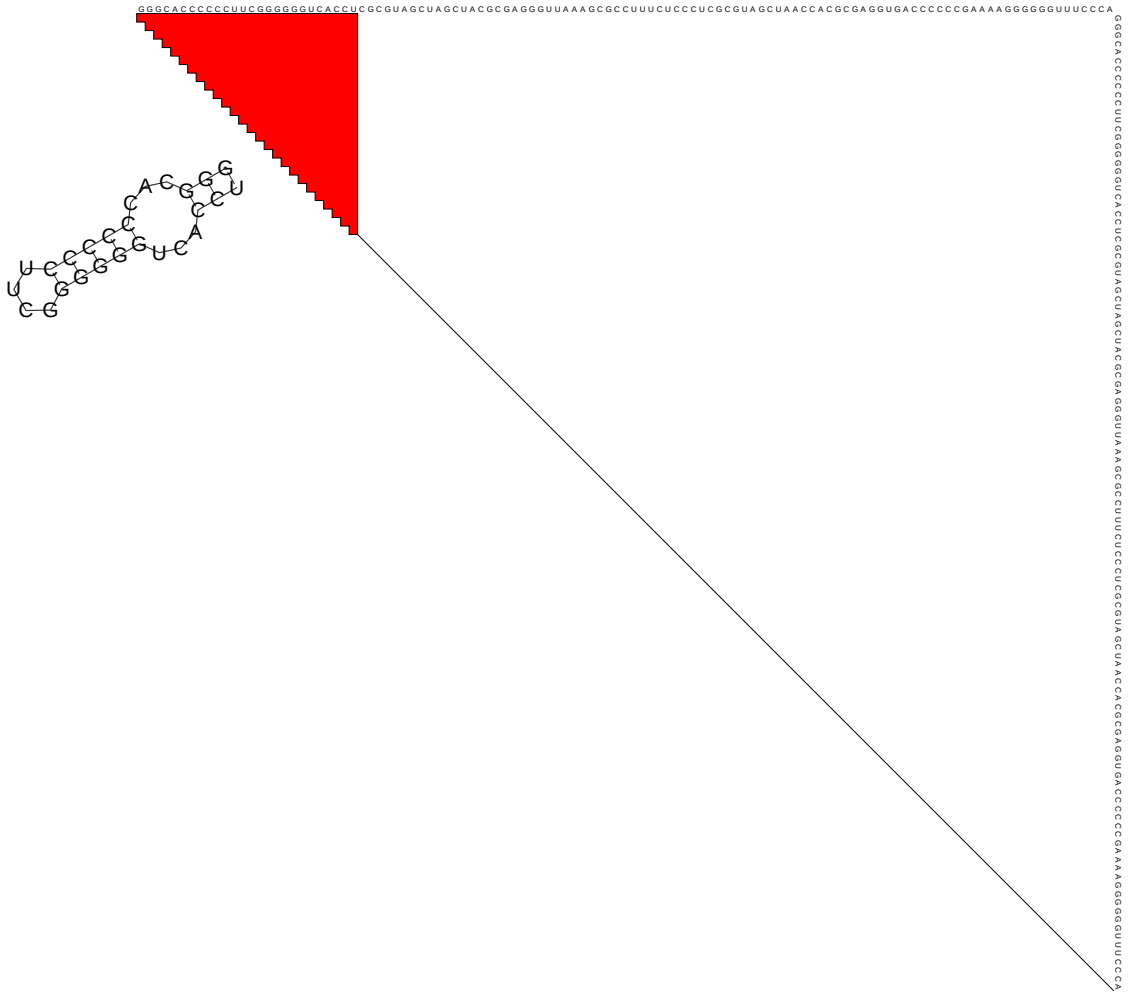
2. Replace further base pairs if it improves energy.

Algorithm

$saddleheight = 0$
$Front\ F = initial\ structure$
$while(F != mfE)$
    increment saddleheight
    for x in extremes
        calculate saddle point S between x and F
        $if(E(S) - E(F) < saddleheight)$
            extend F by x
            remove all x in extremes that conflict with F

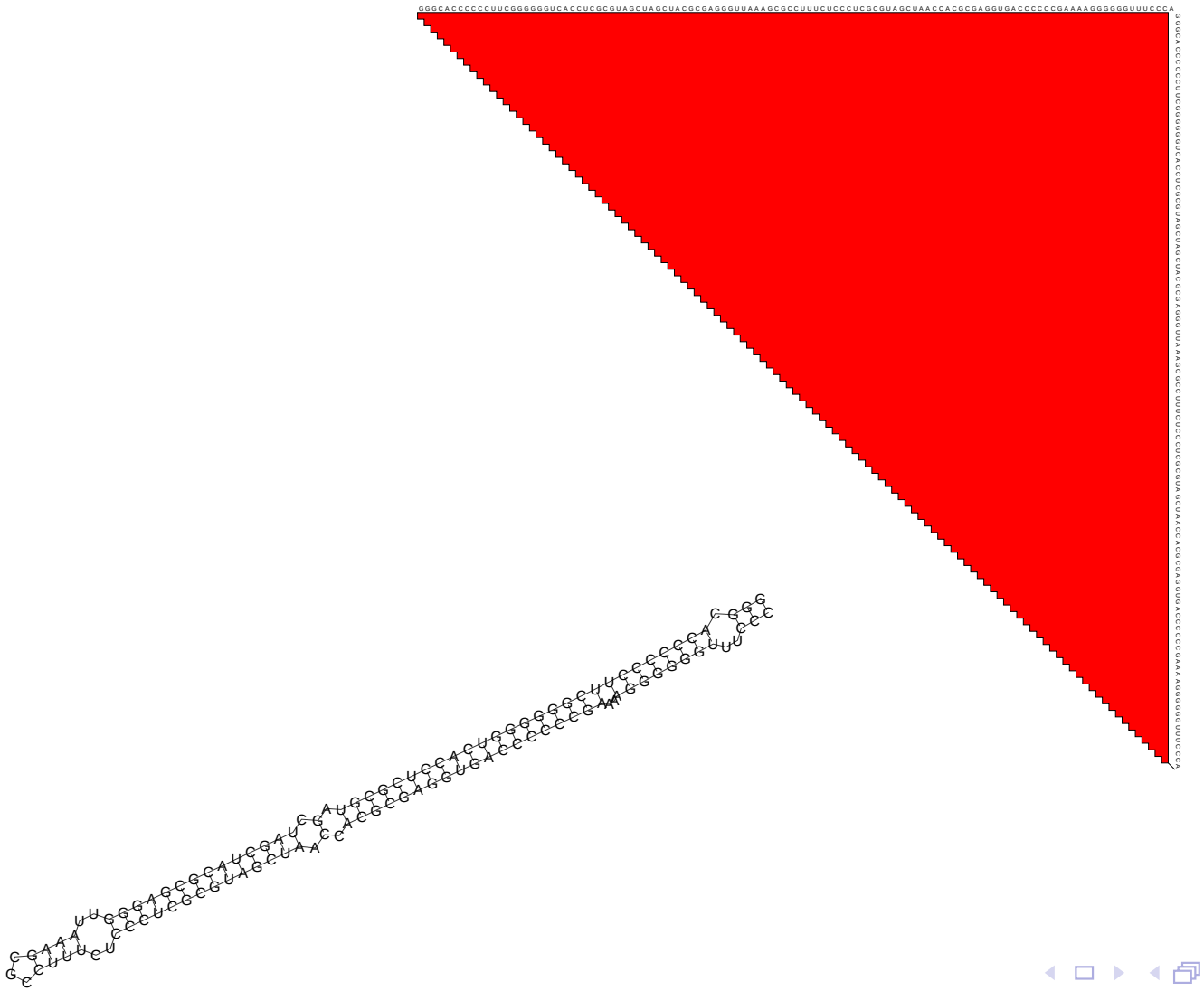
Extend Front

dot.ps



Final Front

dot.ps



Saddle Height

To go from conformation A to B, there are $(A \Delta B)!$ direct paths.

The best path does not have to be direct.

The saddle height is the highest energy of the lowest path:

$$S = min_{P:\ A->B} max_{x \in P} E(x)$$

Heuristic for Saddle Points: Morgan-Higgs(1998)

1. Rank the elements in $B \setminus A$ by conflict
2. For each x in $B \setminus A$
   i) Remove the base pairs in A that conflict with x
   ii) Add x
   iii) Add all other elements in $B \setminus A$ that can be added now
   iv) Record the energies of the traversed states
3. Take the highest recorded energy as saddle energy