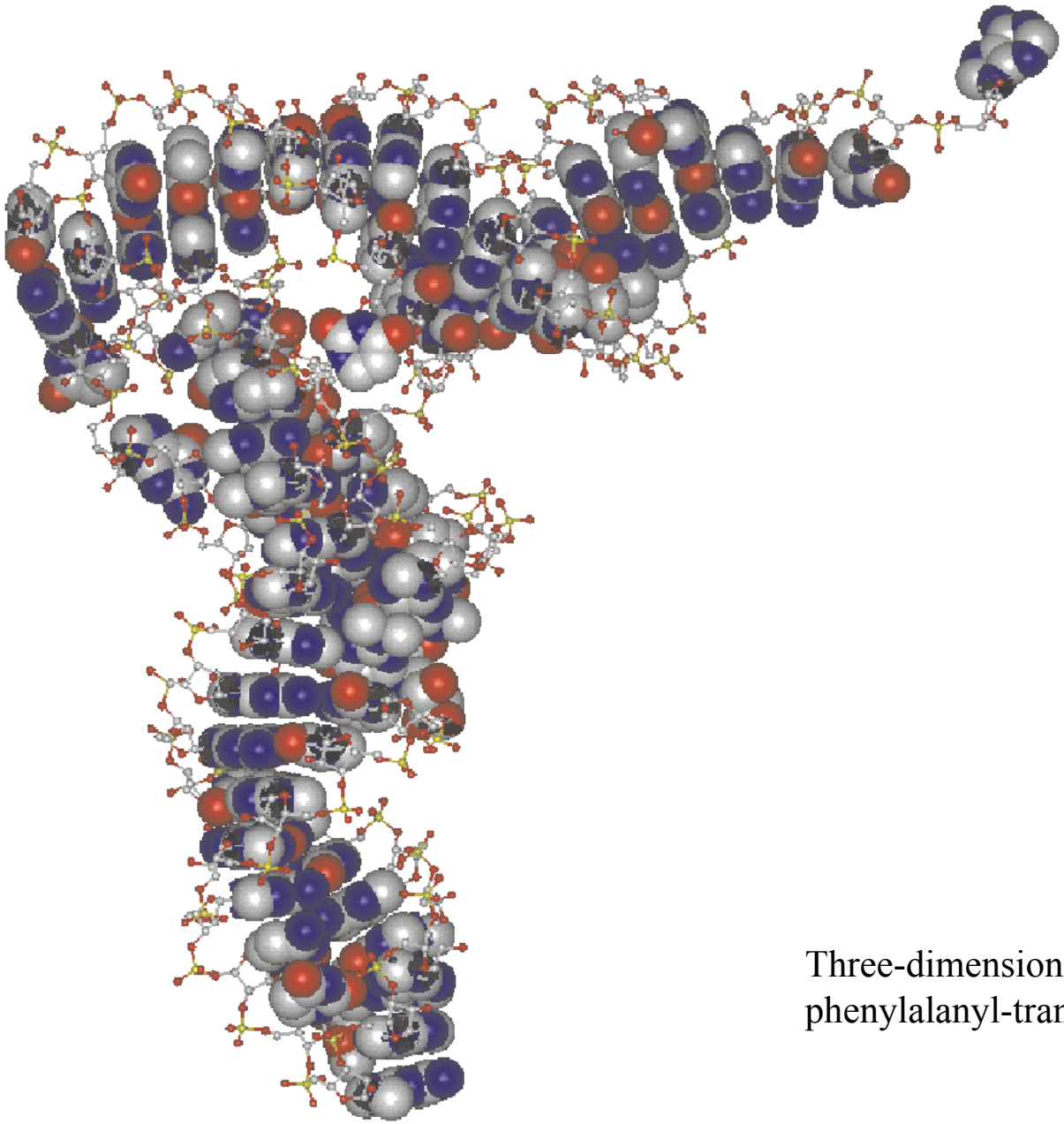
Prediction and Analysis of RNA Secondary Structures

Peter Schuster

Institut für Theoretische Chemie und Molekulare Strukturbiologie der Universität Wien

RNA Secondary Structures in Dijon

Dijon, 24.– 26.06.2002



Three-dimensional structure of
phenylalanyl-transfer-RNA

RNA Secondary Structures and their Properties

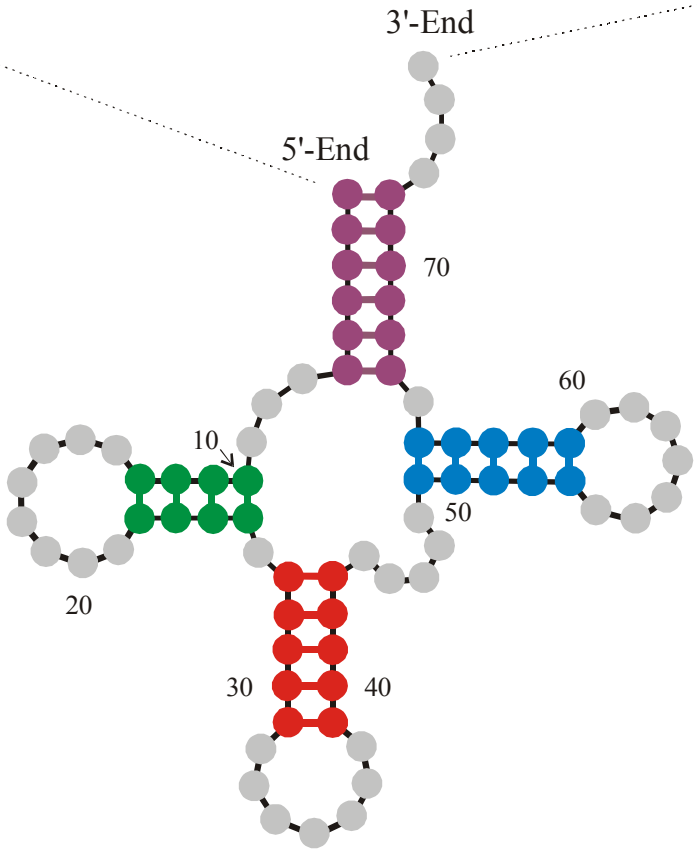
RNA secondary structures are listings of Watson-Crick and GU wobble base pairs, which are free of knots and pseudokots. Secondary structures are **folding intermediates** in the formation of full three-dimensional structures.

D.Thirumalai, N.Lee, S.A.Woodson, and D.K.Klimov.
Annu.Rev.Phys.Chem. **52**:751-762 (2001)

Sequence

5'-End GCGGAUUUAGCUCAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCCUGUGTPCGAUCCACAGAAUUCGCACCA 3'-End

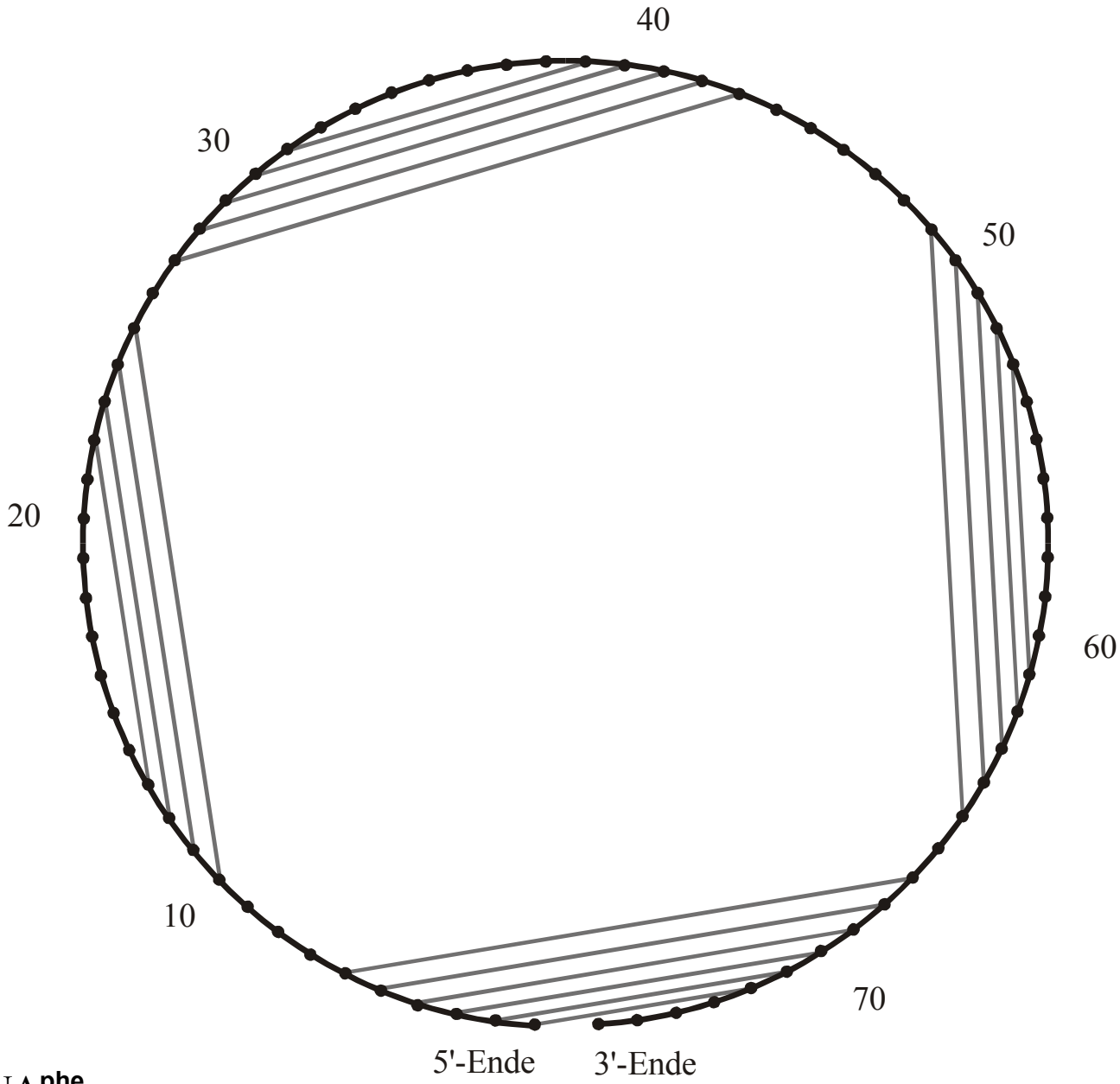
Secondary Structure



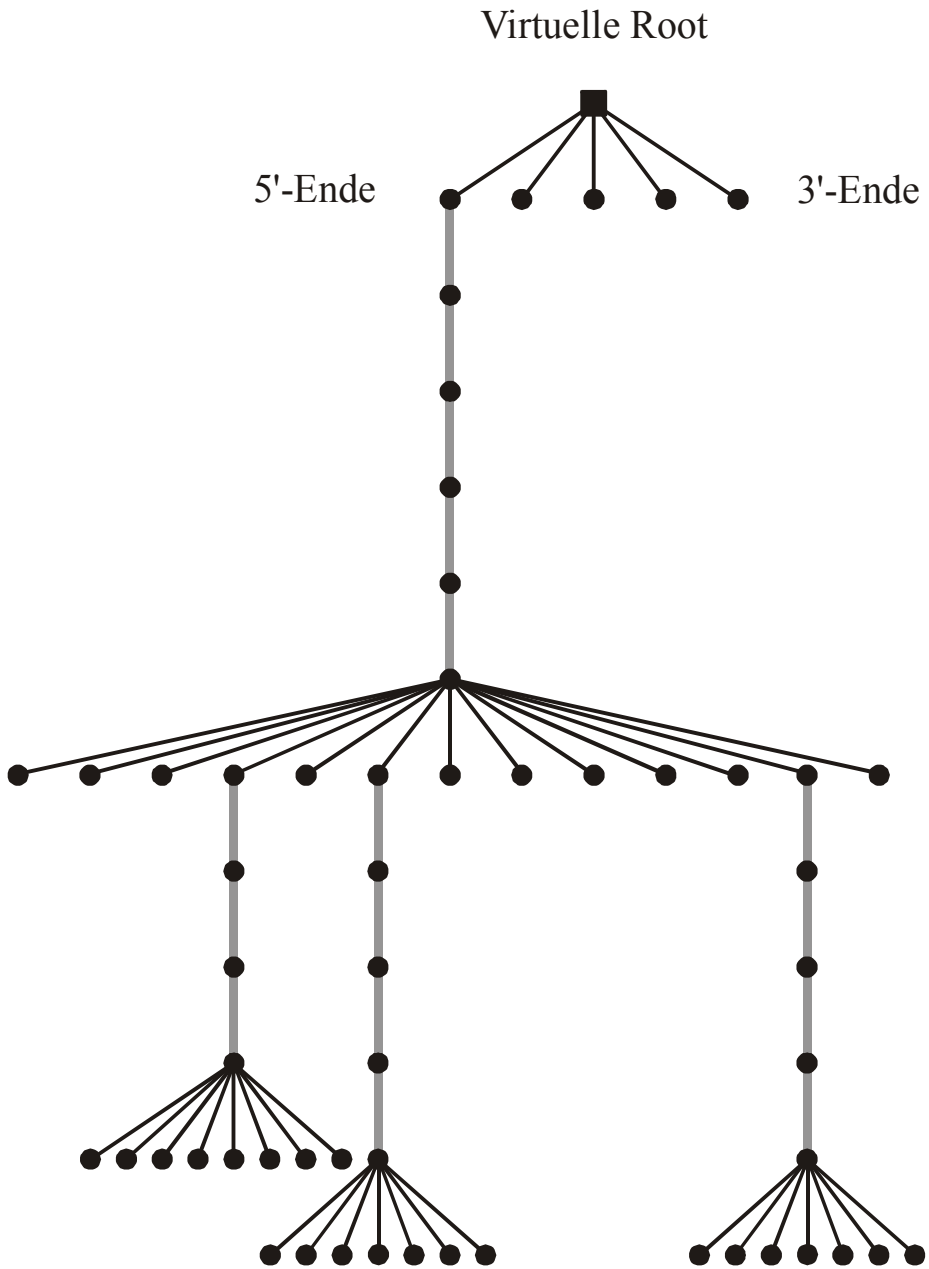
Symbolic Notation

5'-End ((((((···((((·····))))·((((·····))))).....((((·····))))).))))···· 3'-End

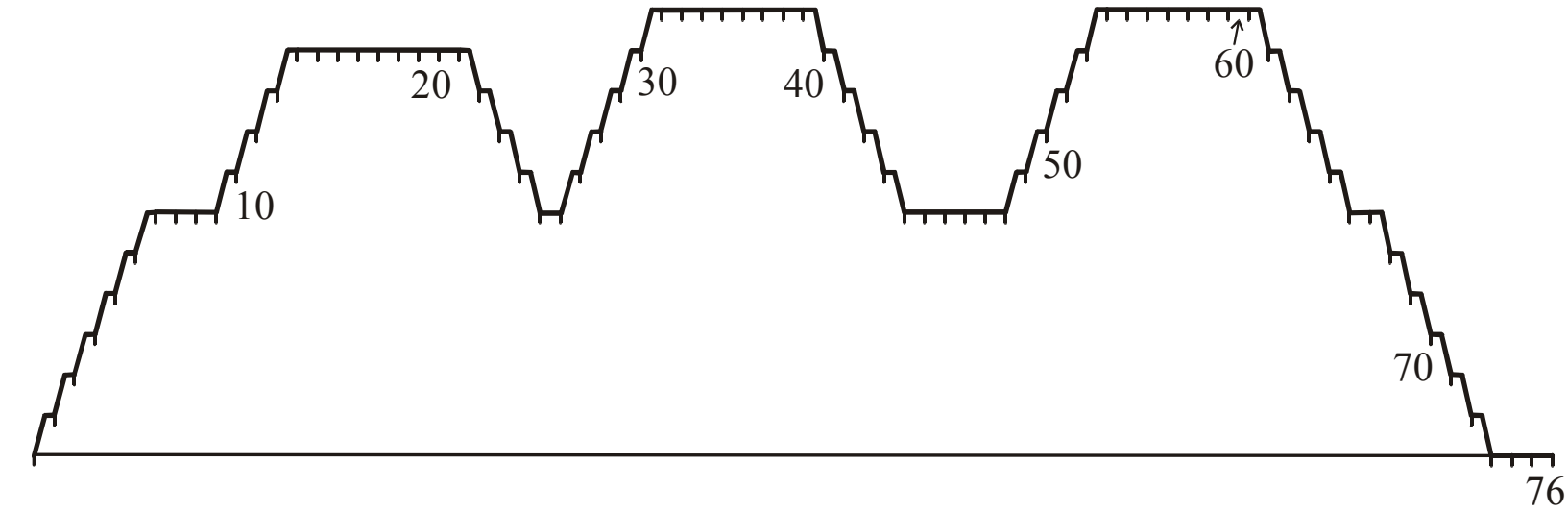
Definition and formation of the secondary structure of phenylalanyl-tRNA



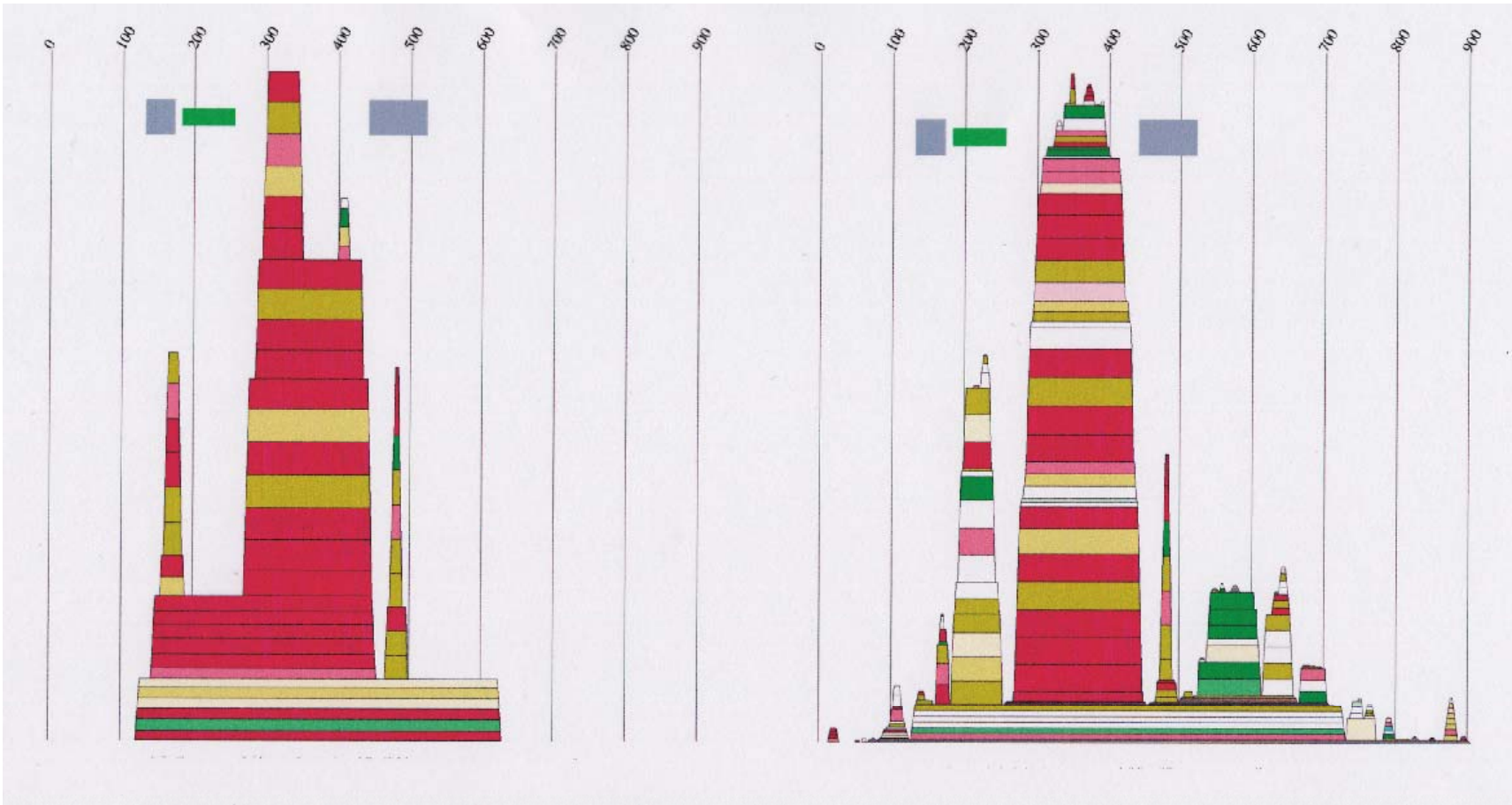
Circle representation of tRNA^{phe}



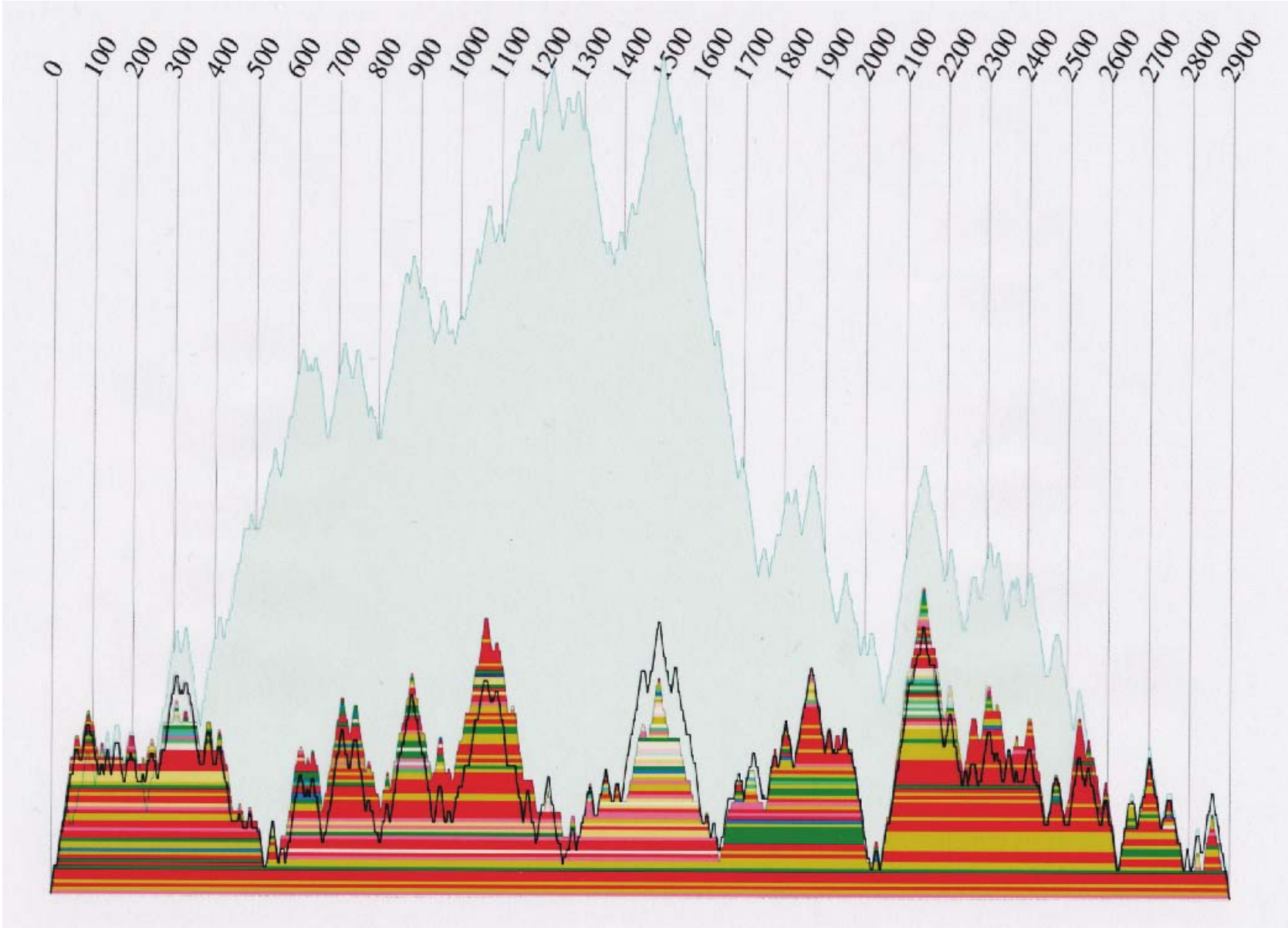
Tree representation of $tRNA^{phe}$



Mountain representation of tRNA^{phe}

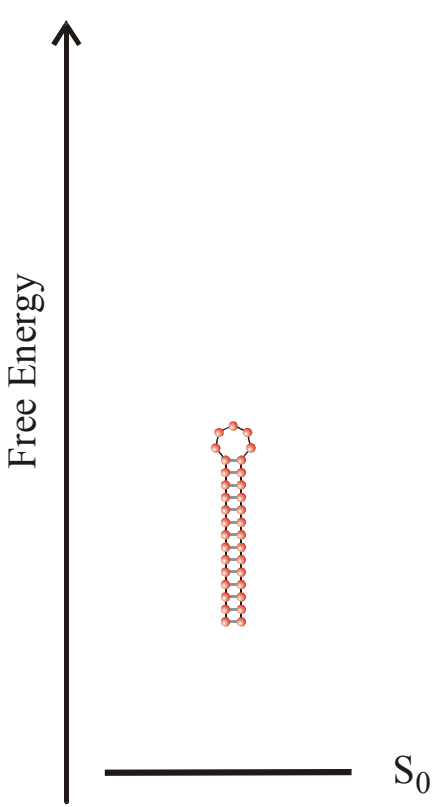


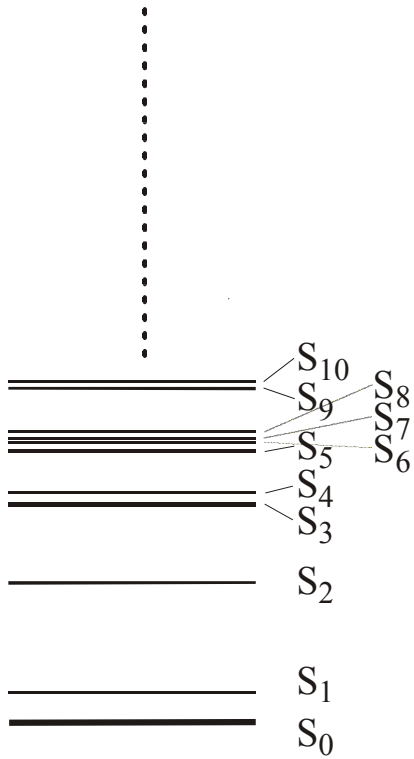
Mountain representation used in structure prediction of medium size RNA molecules

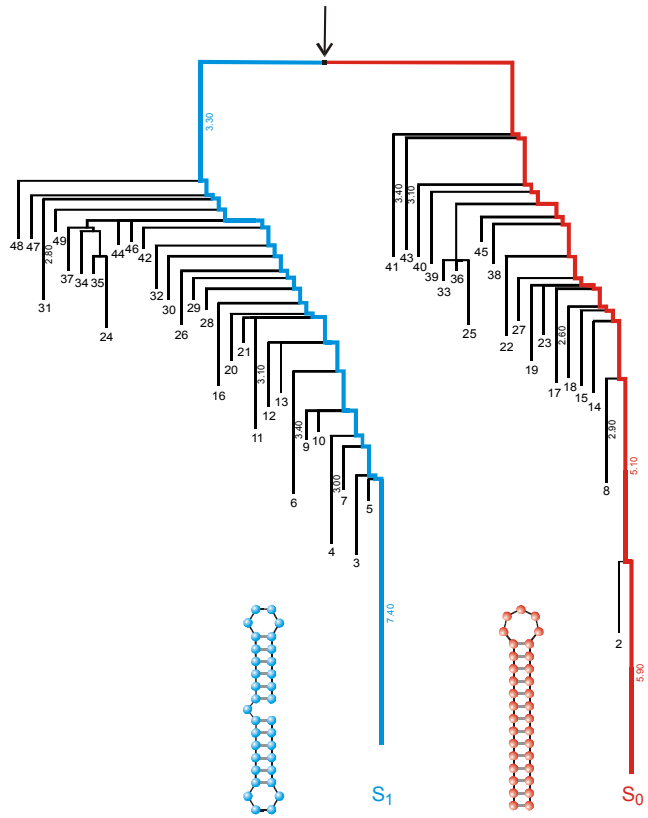


Mountain representation used in structure prediction of large RNA molecules

T = 0 K , t Á ¸ T > 0 K , t Á ¸ T > 0 K , t finite







Minimum Free Energy Structure Suboptimal Structures Kinetic Structures

Different notions of RNA structure

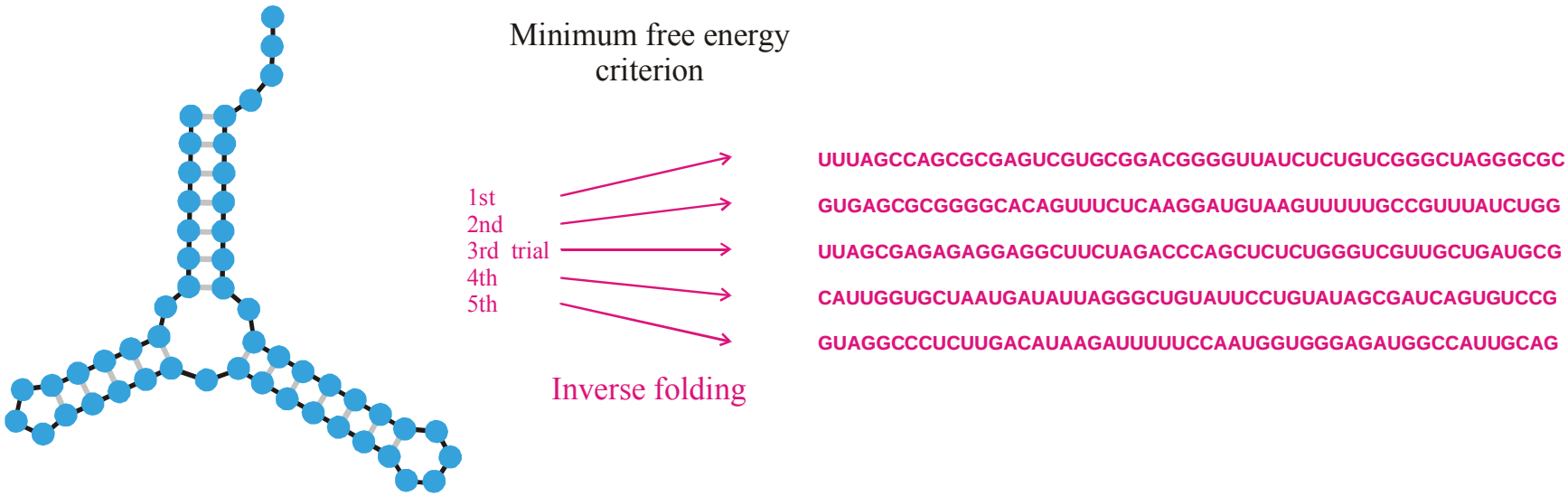
RNA Minimum Free Energy Structures

Efficient algorithms based on dynamical programming are available for computation of secondary structures for given sequences. Inverse folding algorithms compute sequences for given secondary structures.

M.Zuker and P.Stiegler. *Nucleic Acids Res*. **9**:133-148 (1981)

Vienna RNA Package: http:www.tbi.univie.ac.at (includes **inverse folding**, **suboptimal structures**, **kinetic folding**, etc.)

I.L.Hofacker, W. Fontana, P.F.Stadler, L.S.Bonhoeffer, M.Tacker, and P. Schuster. *Mh.Chem*. **125**:167-188 (1994)



The inverse folding algorithm searches for sequences that form a given RNA secondary structure under the minimum free energy criterion.

Criterion of
Minimum Free Energy

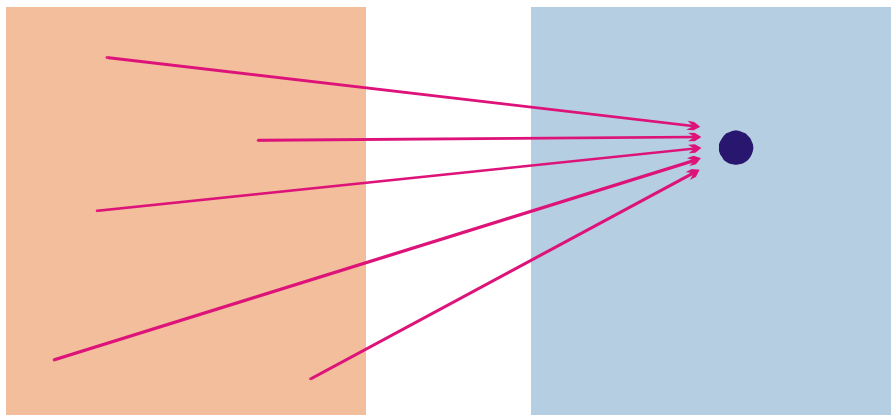
UUUAGCCAGCGCGAGUCGUGCGGACGGGGUUAUCUCUGUCGGGCUAGGGCGC

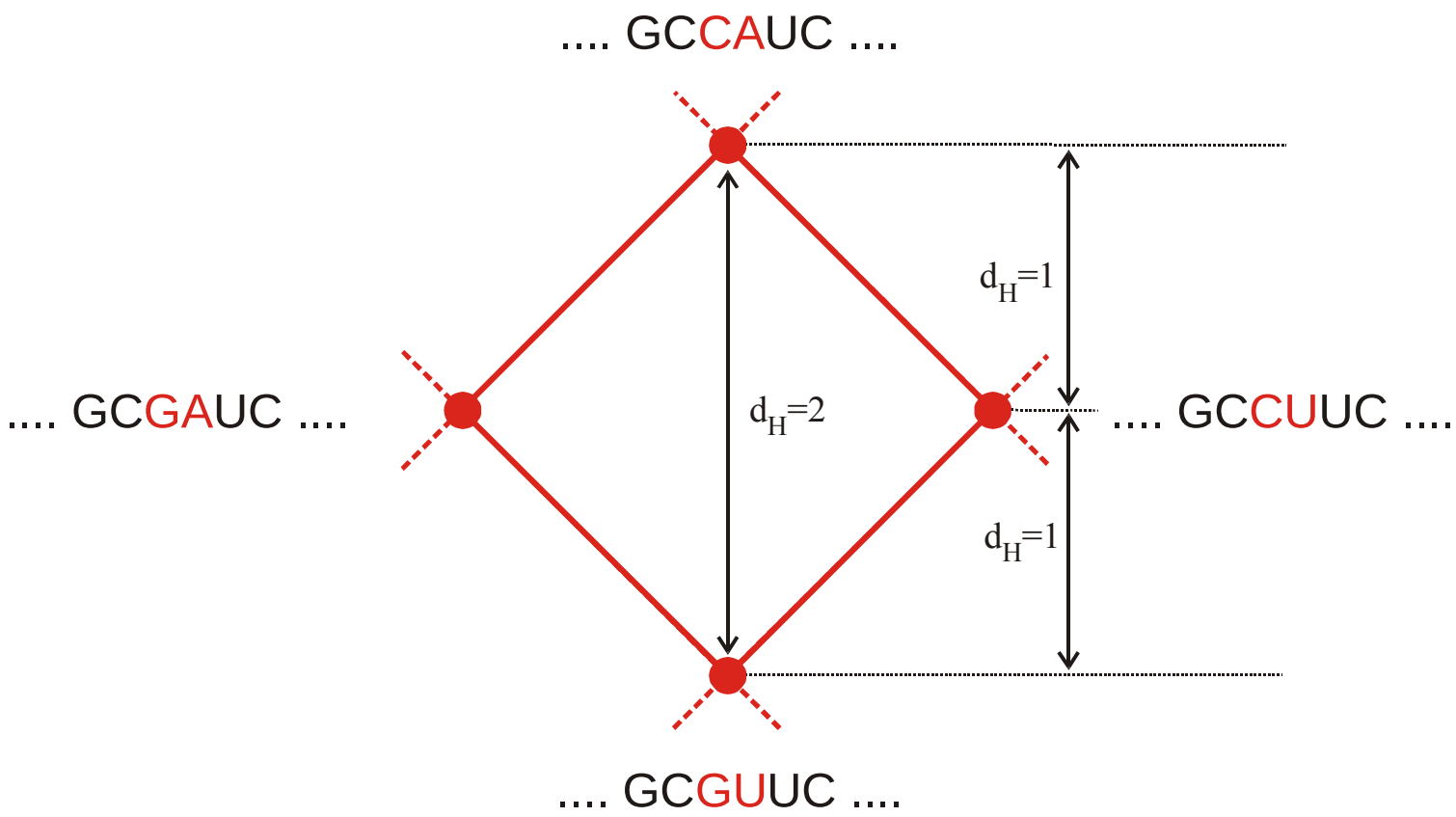
GUGAGCGCGGGGCACAGUUUCUCAAGGAUGUAAGUUUUUGCCGUUUAUCUGG

UUAGCGAGAGAGGAGGCUUCUAGACCCAGCUCUCUGGGUCGUUGCUGAUGCG

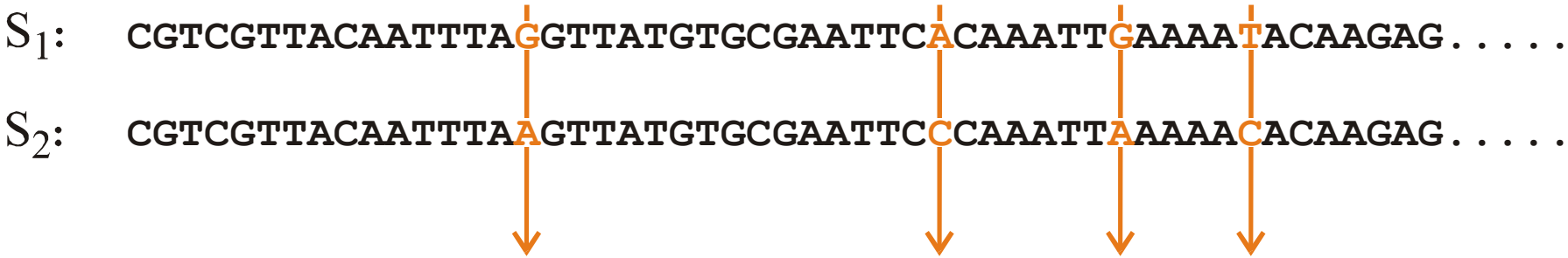
CAUUGGUGCUAAUGAUAUUAGGGCUGUAUUCCUGUAUAGCGAUCAGUGUCCG

GUAGGCCCUCUUGACAUAAGAUUUUUCCAAUGGUGGGAGAUGGCCAUUGCAG





Point mutations as moves in sequence space



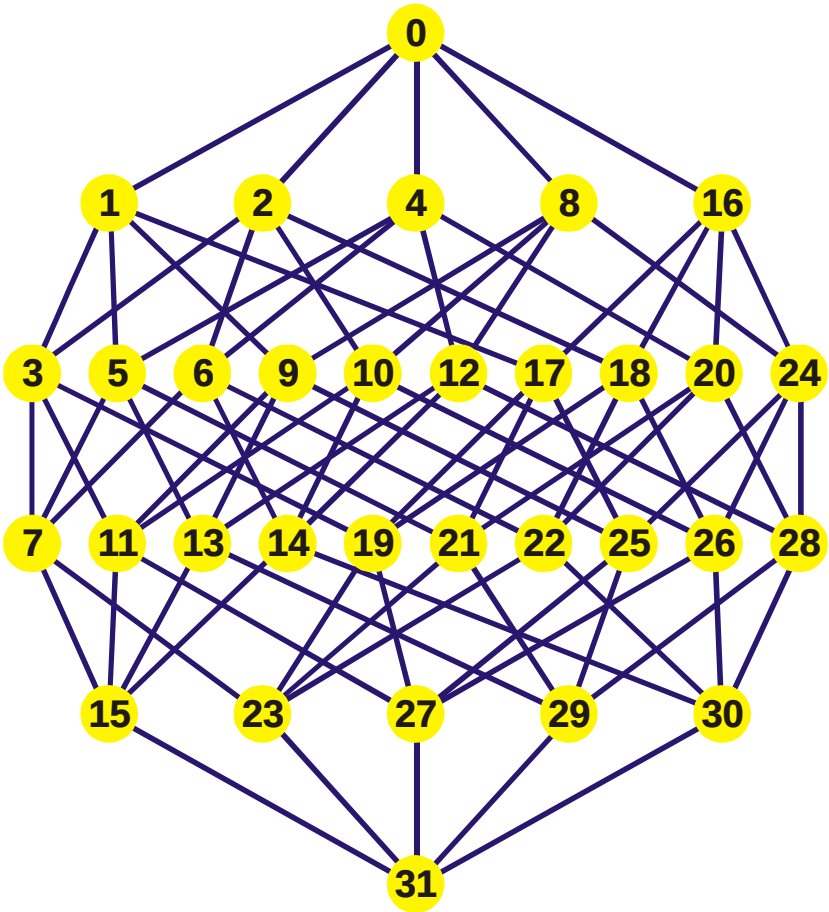
Hamming distance $d_H(S_1,S_2) = 4$

(i) $d_H(S_1,S_1) = 0$

(ii) $d_H(S_1,S_2) = d_H(S_2,S_1)$

(iii) $d_H(S_1,S_3) \P d_H(S_1,S_2) + d_H(S_2,S_3)$

The Hamming distance induces a metric in sequence space



Mutant class

0

1

2

3

4

5

Binary sequences are encoded by their decimal equivalents:

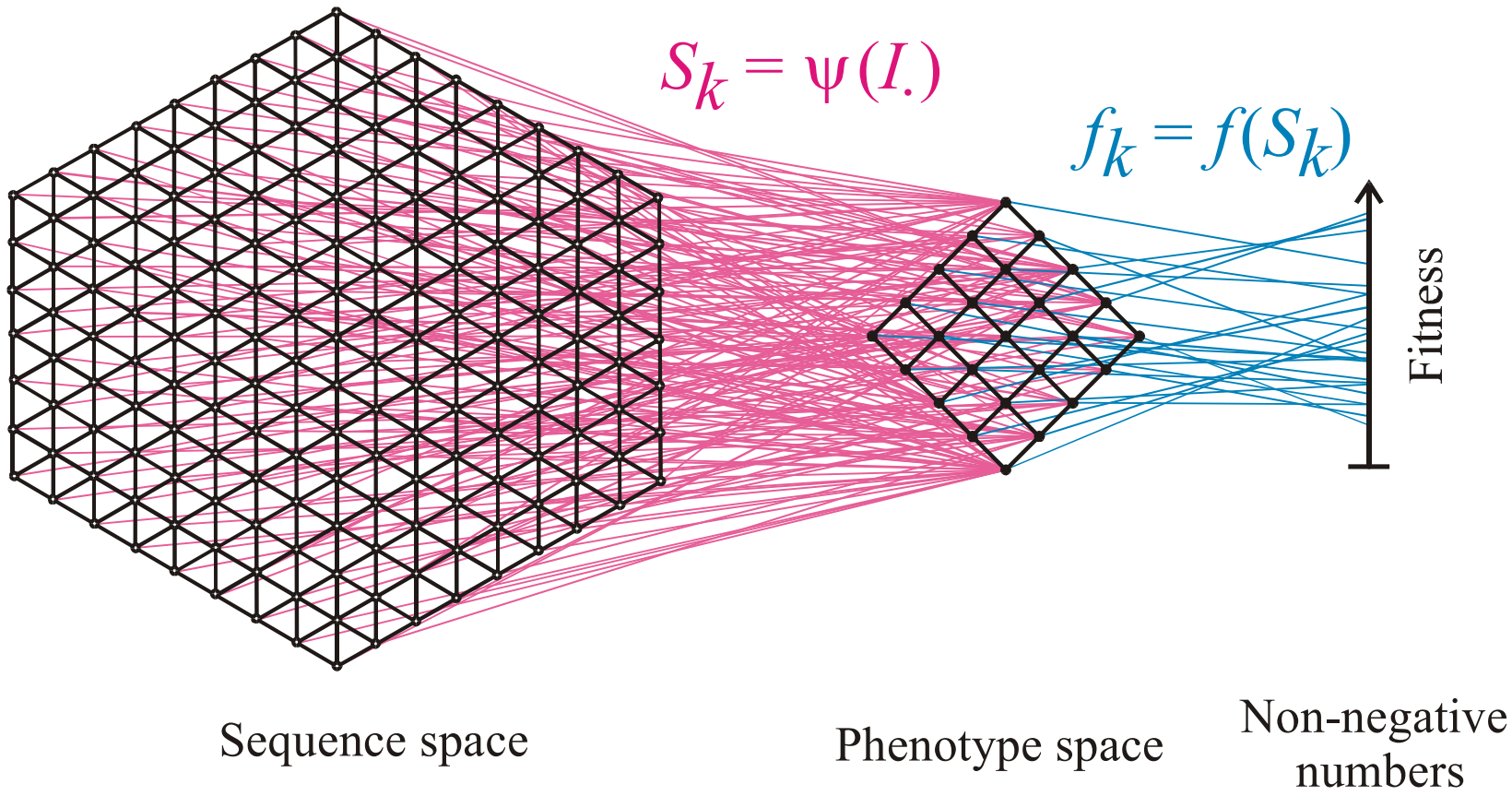
C = 0 and G = 1, for example,

"0" ≡ 00000 = CCCCC,

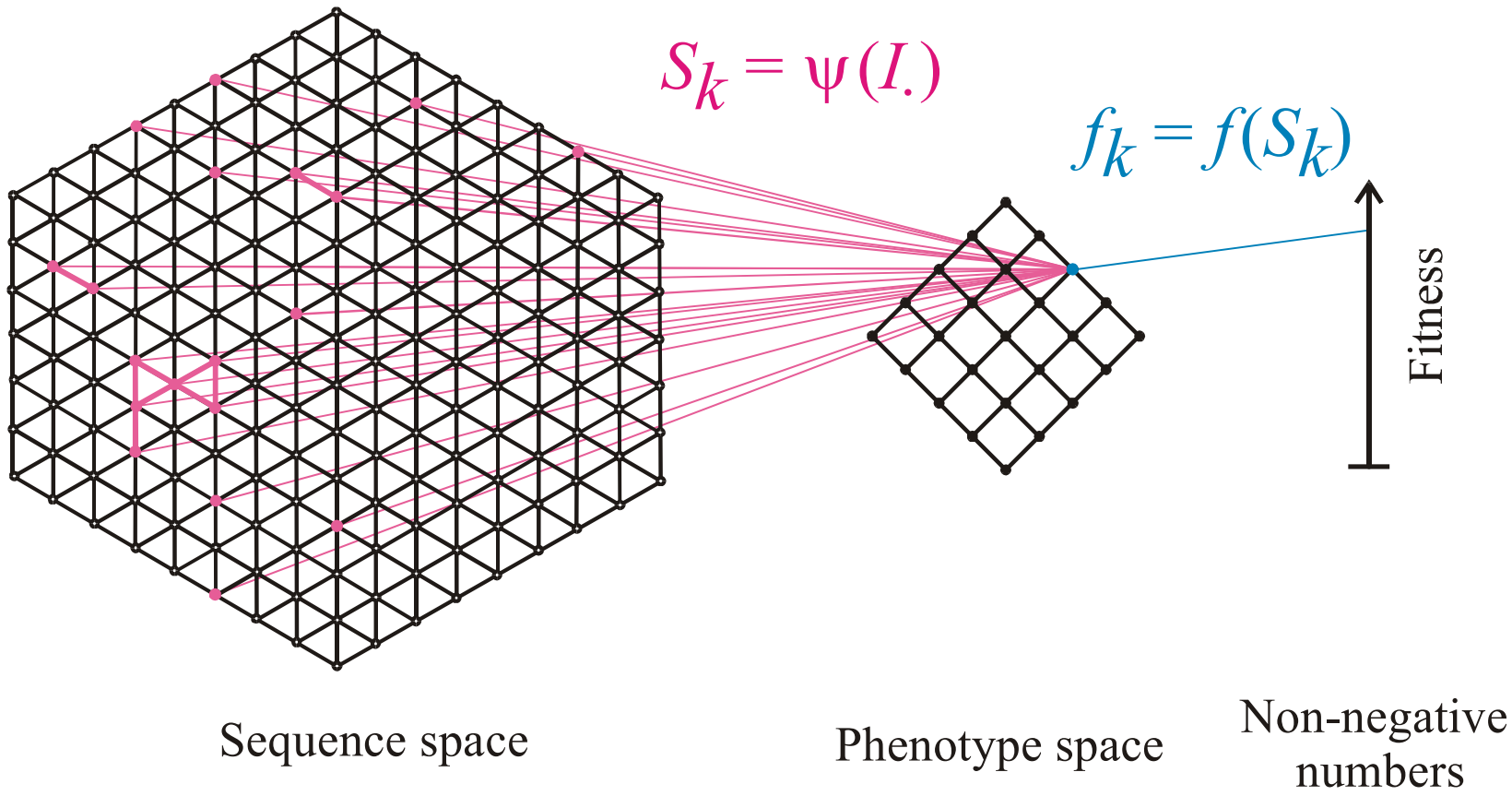
"14" ≡ 01110 = CGGGC,

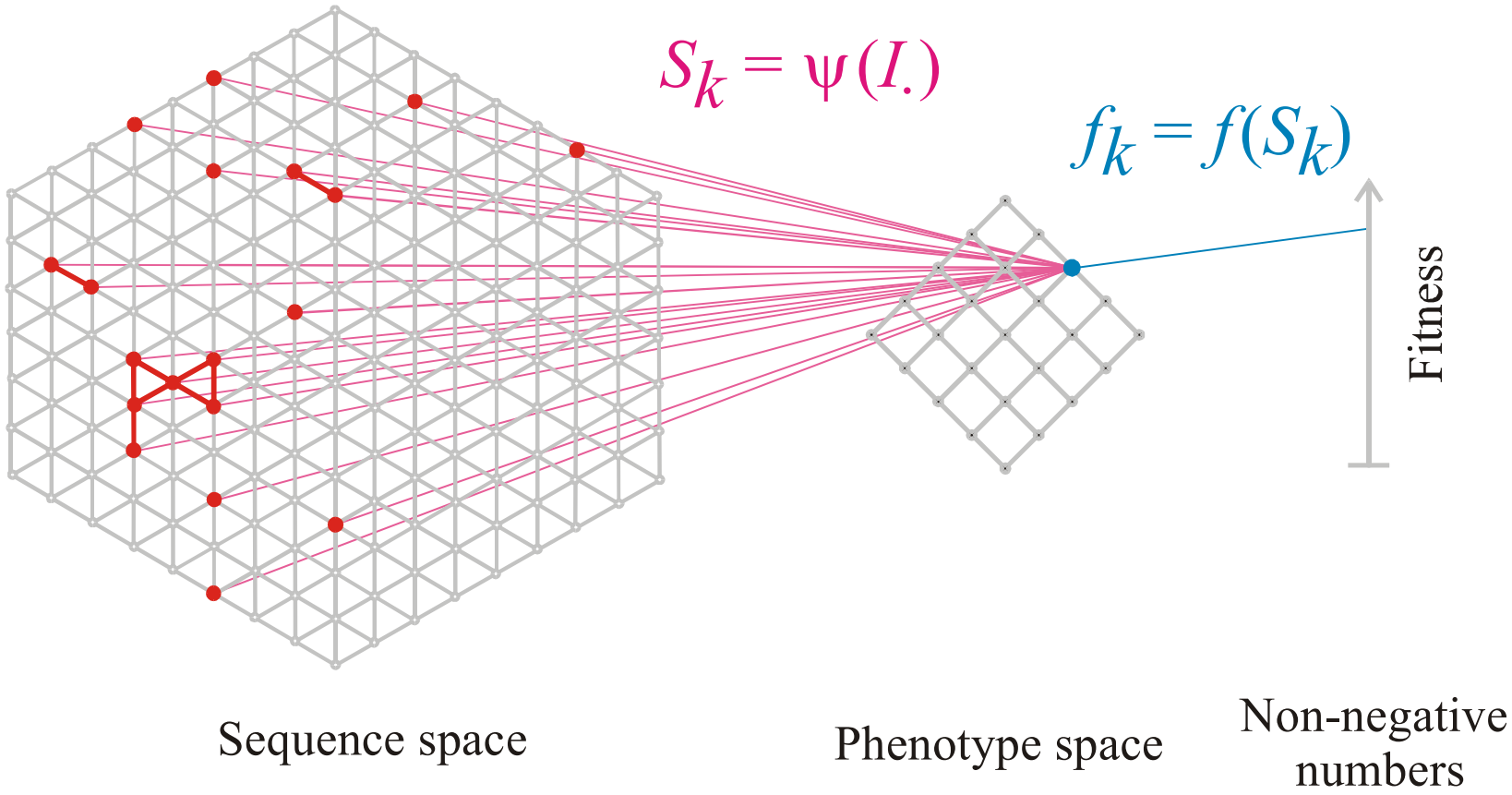
"29" ≡ 11101 = GGGCG, etc.

Sequence space of binary sequences of chain lenght n=5



Mapping from sequence space into phenotype space and into fitness values


$S_k = \psi(I_\cdot)$
$f_k = f(S_k)$
Fitness
Sequence space
Phenotype space
Non-negative numbers

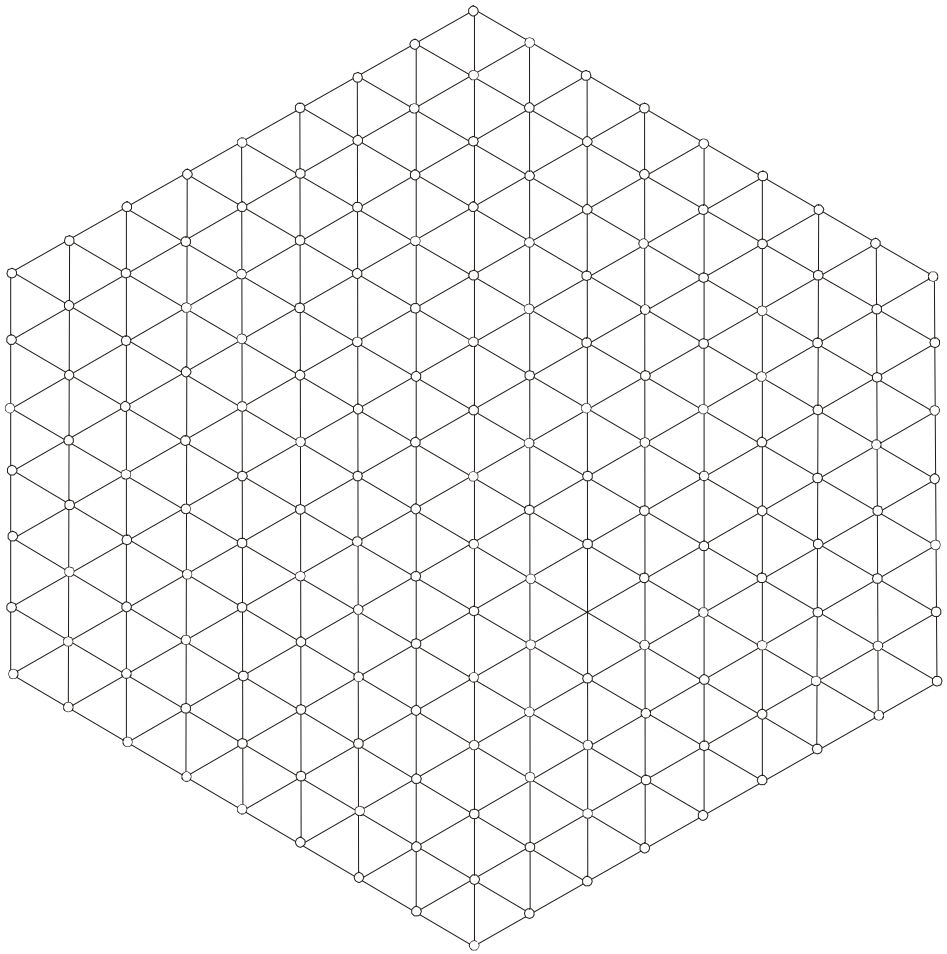

$S_k = \psi(I_\cdot)$
$f_k = f(S_k)$
Fitness
Sequence space
Phenotype space
Non-negative numbers

Neutral networks of small RNA molecules can be computed by exhaustive folding of complete sequence spaces, i.e. all RNA sequences of a given chain length. This number, $\mathbf{N=4^n}$, becomes very large with increasing length, and is prohibitive for numerical computations.

Neutral networks can be modelled by **random graphs** in sequence space. In this approach, nodes are inserted randomly into sequence space until the size of the pre-image, i.e. the number of neutral sequences, matches the neutral network to be studied.

Step 00

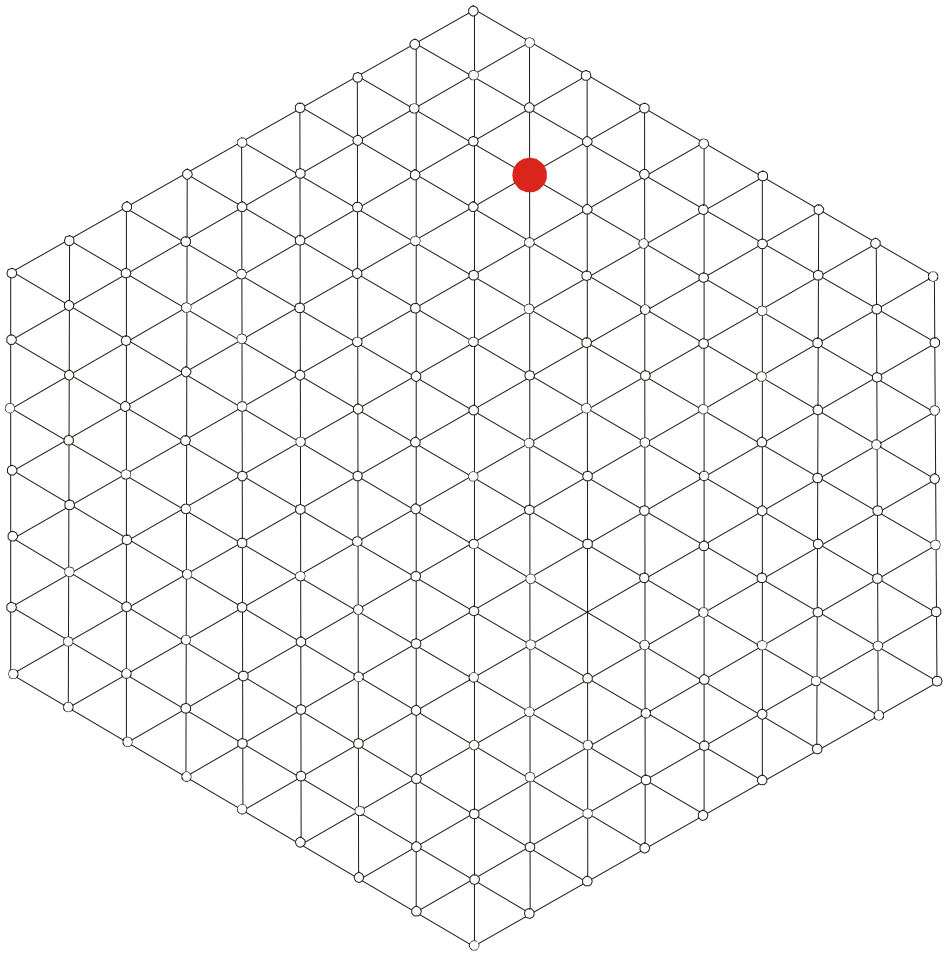
Sketch of sequence space



Random graph approach to neutral networks

Step 01

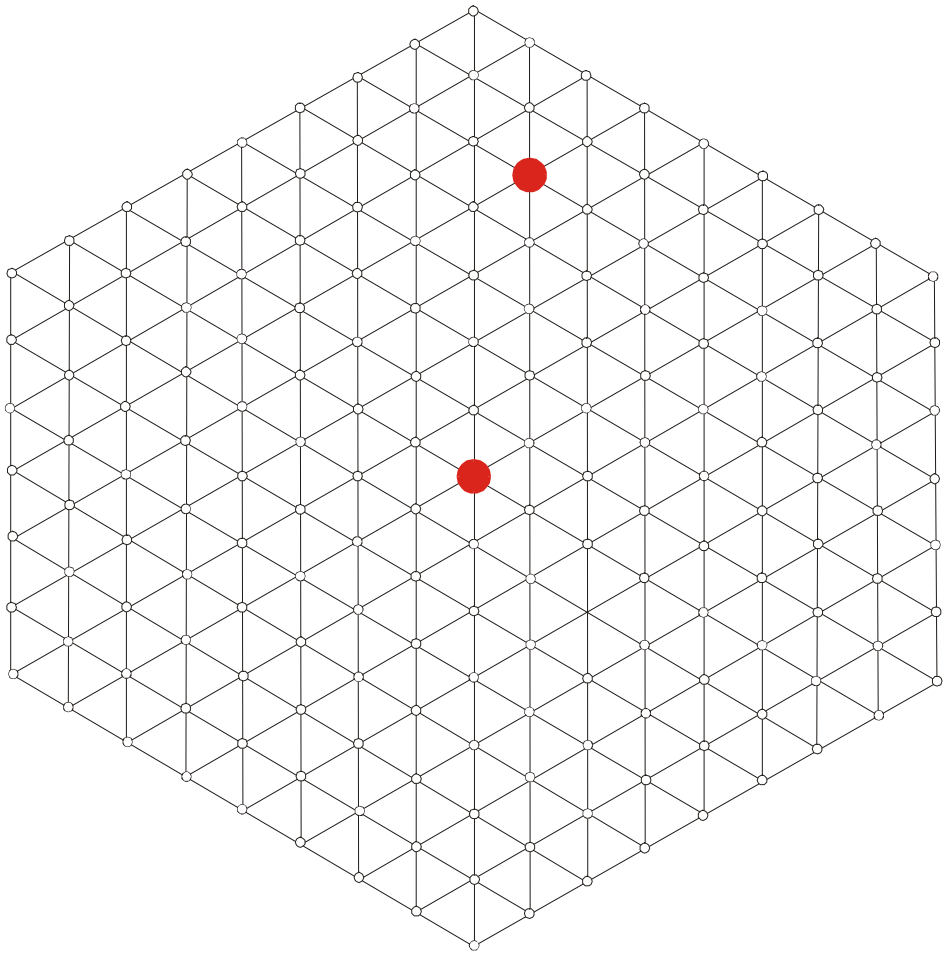
Sketch of sequence space



Random graph approach to neutral networks

Step 02

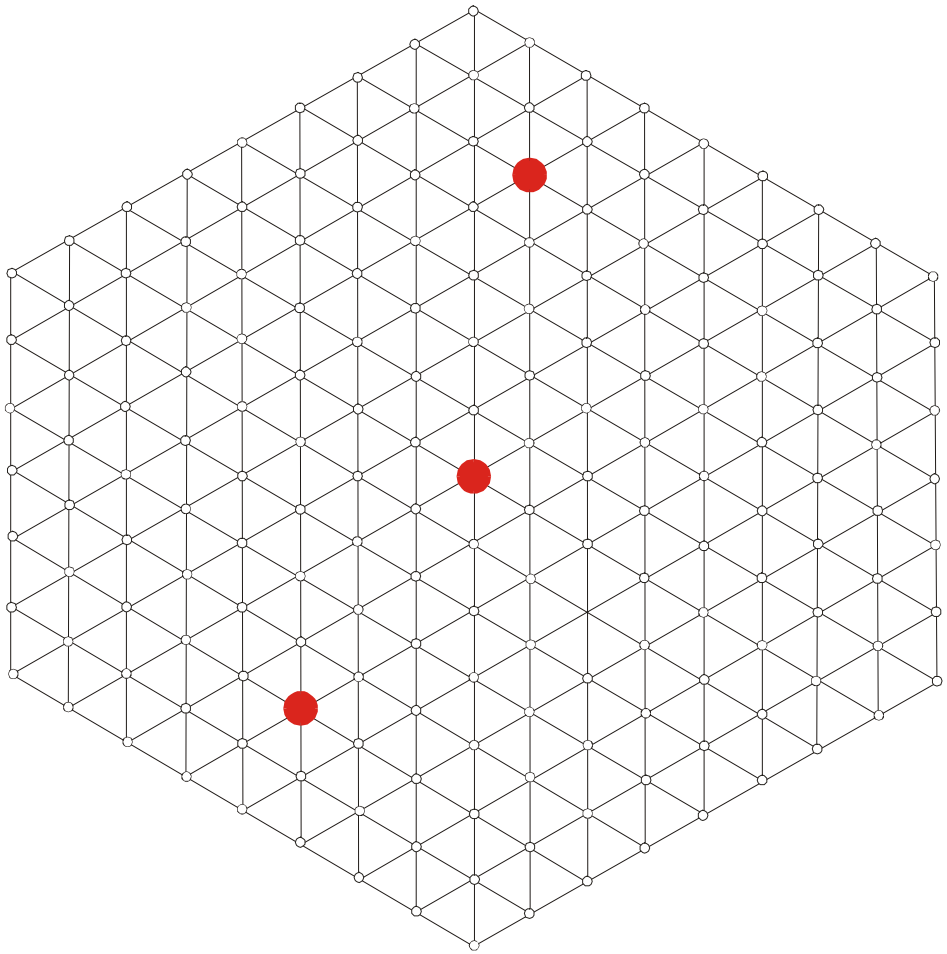
Sketch of sequence space



Random graph approach to neutral networks

Step 03

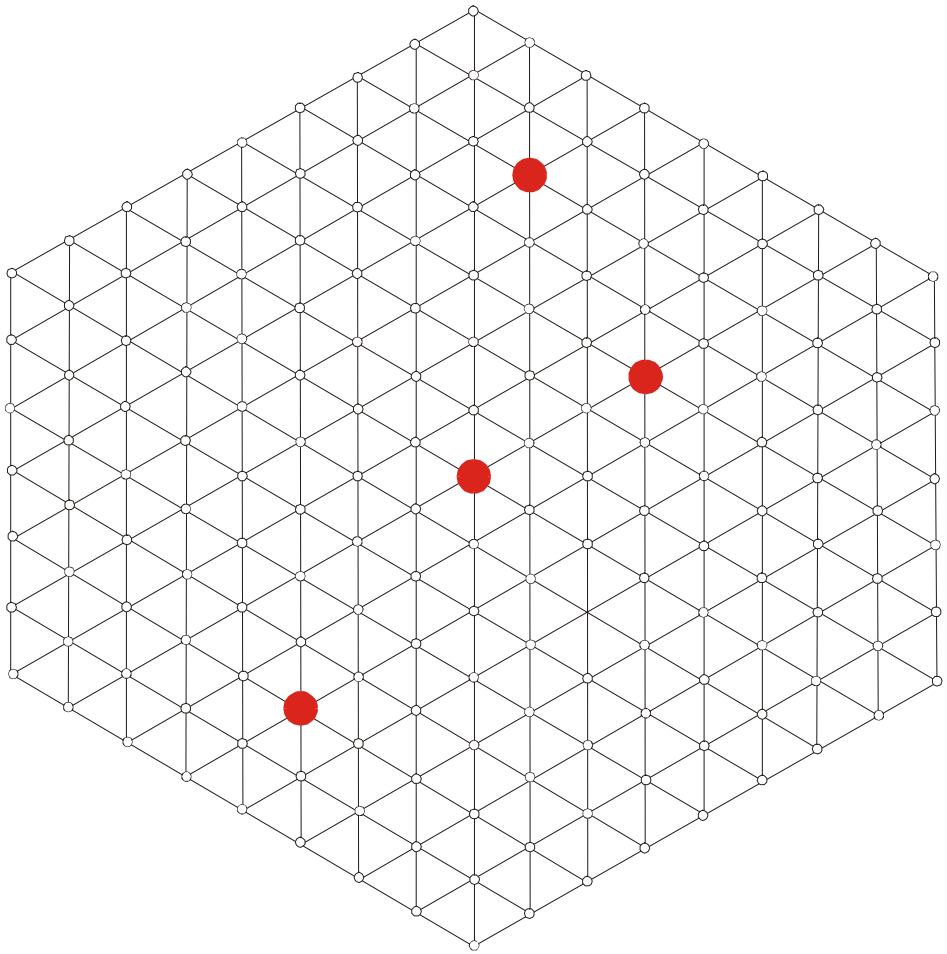
Sketch of sequence space



Random graph approach to neutral networks

Step 04

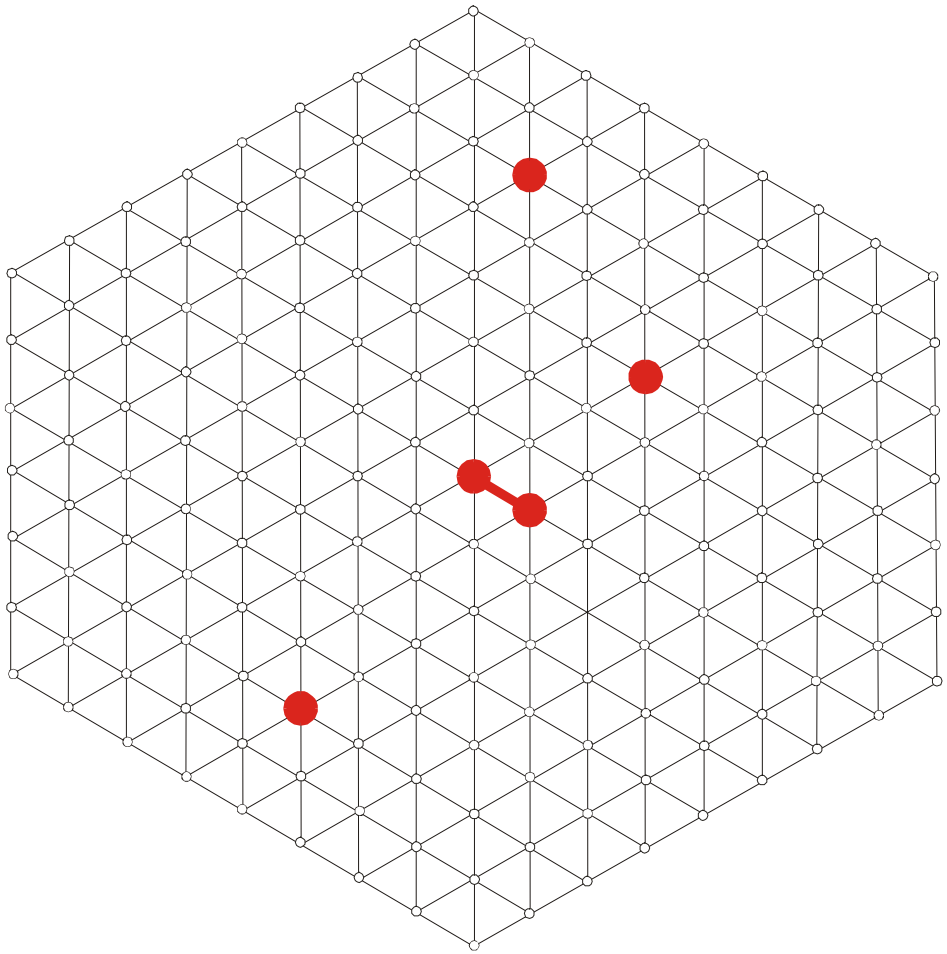
Sketch of sequence space



Random graph approach to neutral networks

Step 05

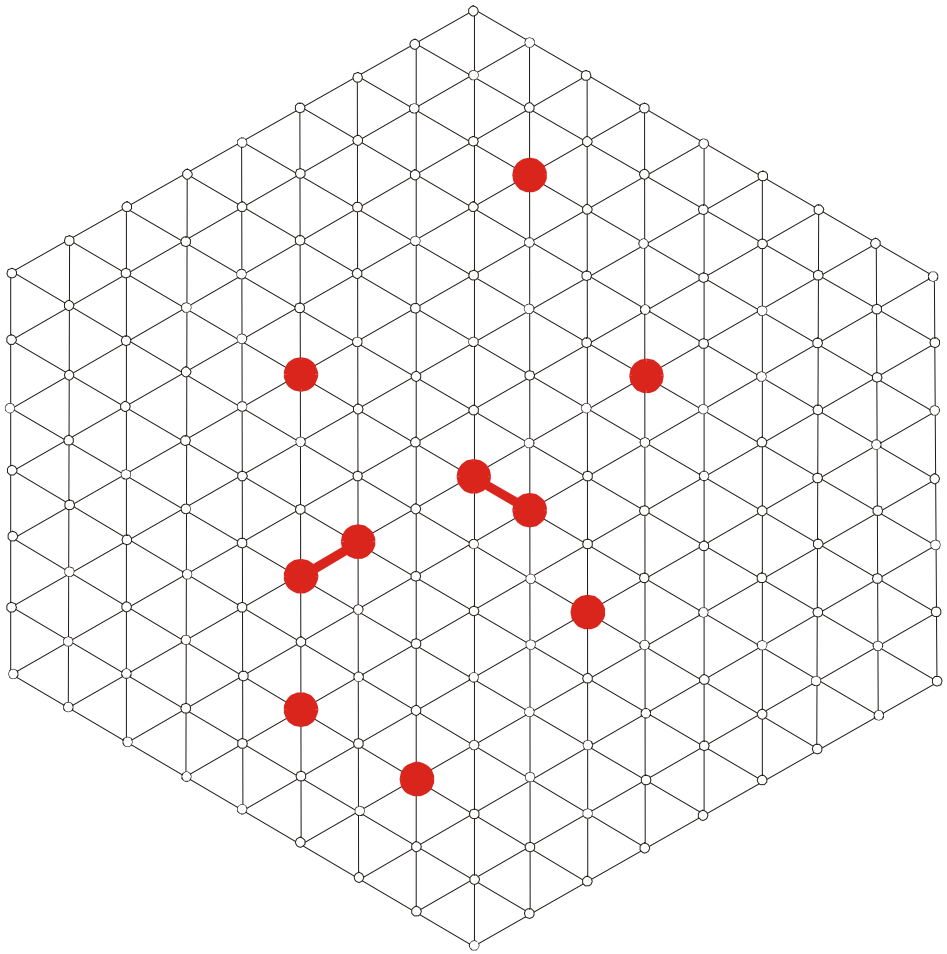
Sketch of sequence space



Random graph approach to neutral networks

Step 10

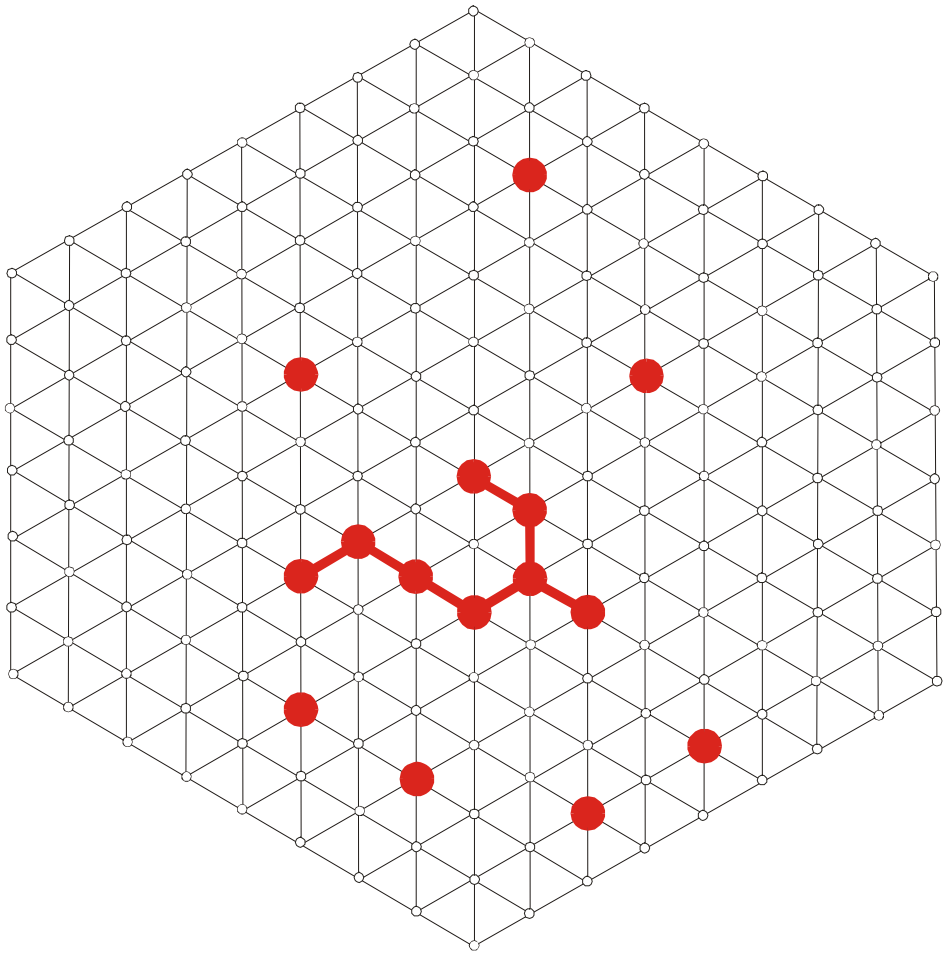
Sketch of sequence space



Random graph approach to neutral networks

Step 15

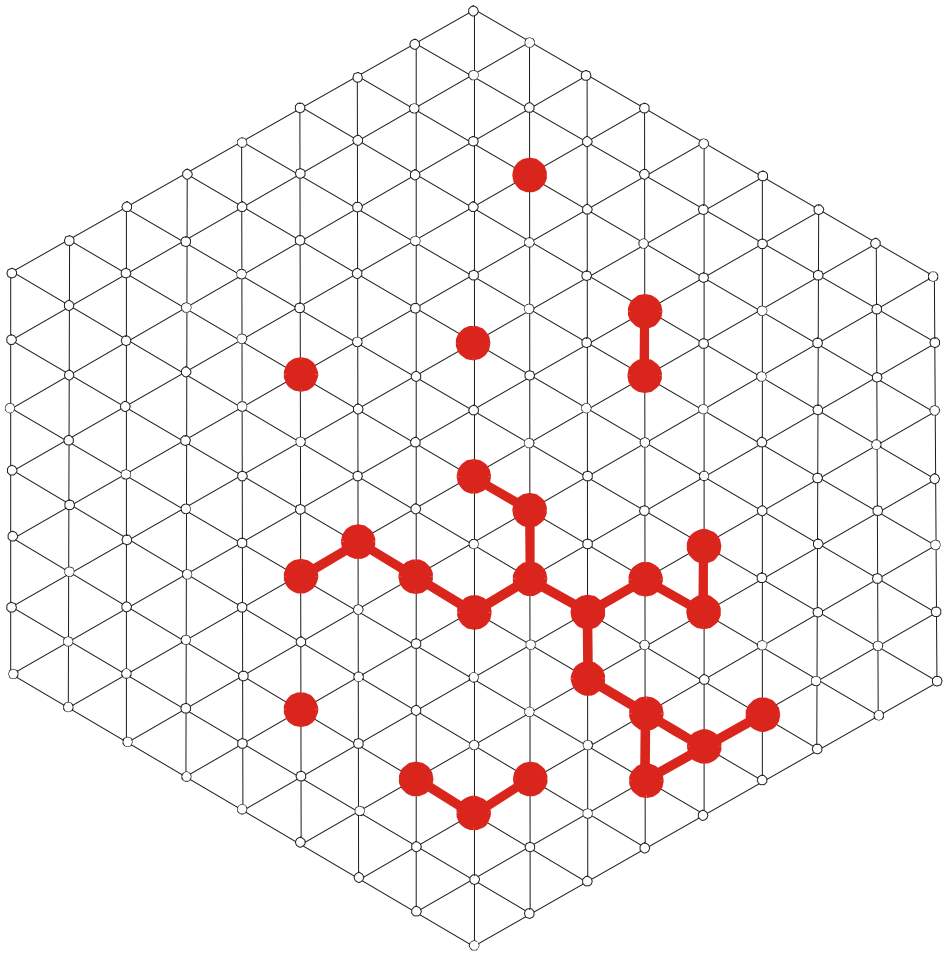
Sketch of sequence space



Random graph approach to neutral networks

Step 25

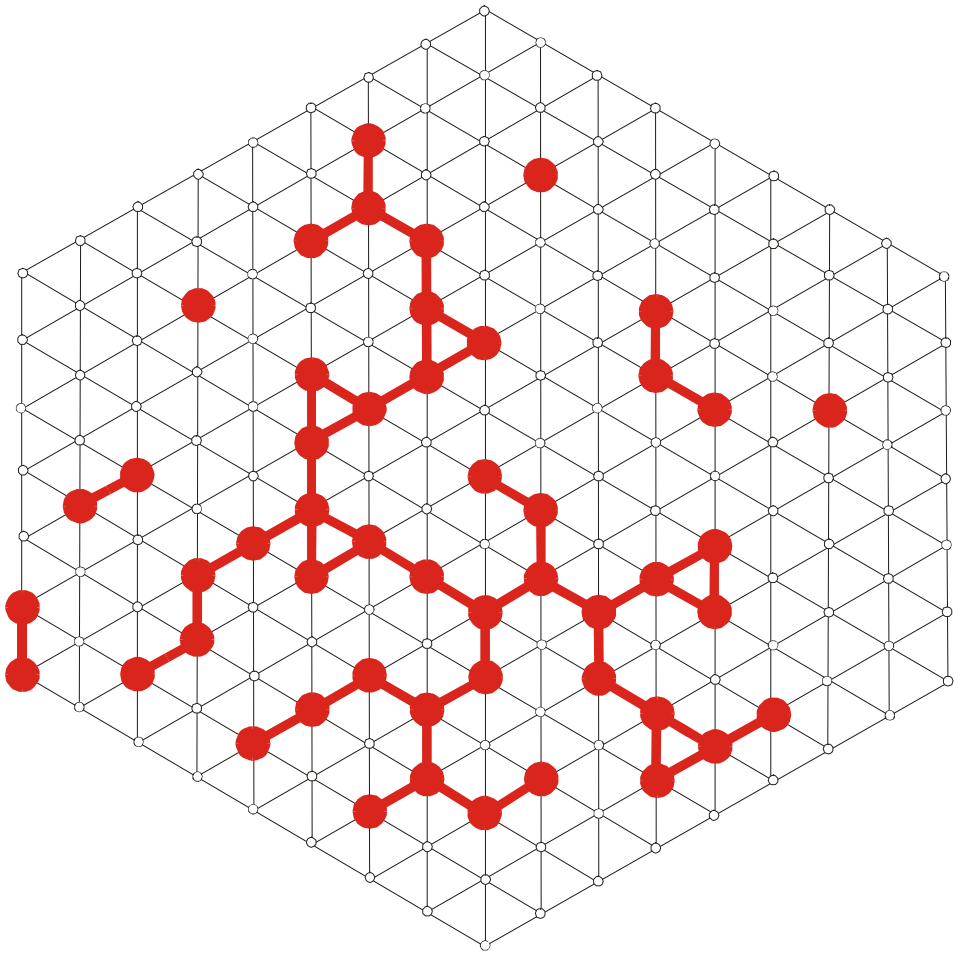
Sketch of sequence space



Random graph approach to neutral networks

Step 50

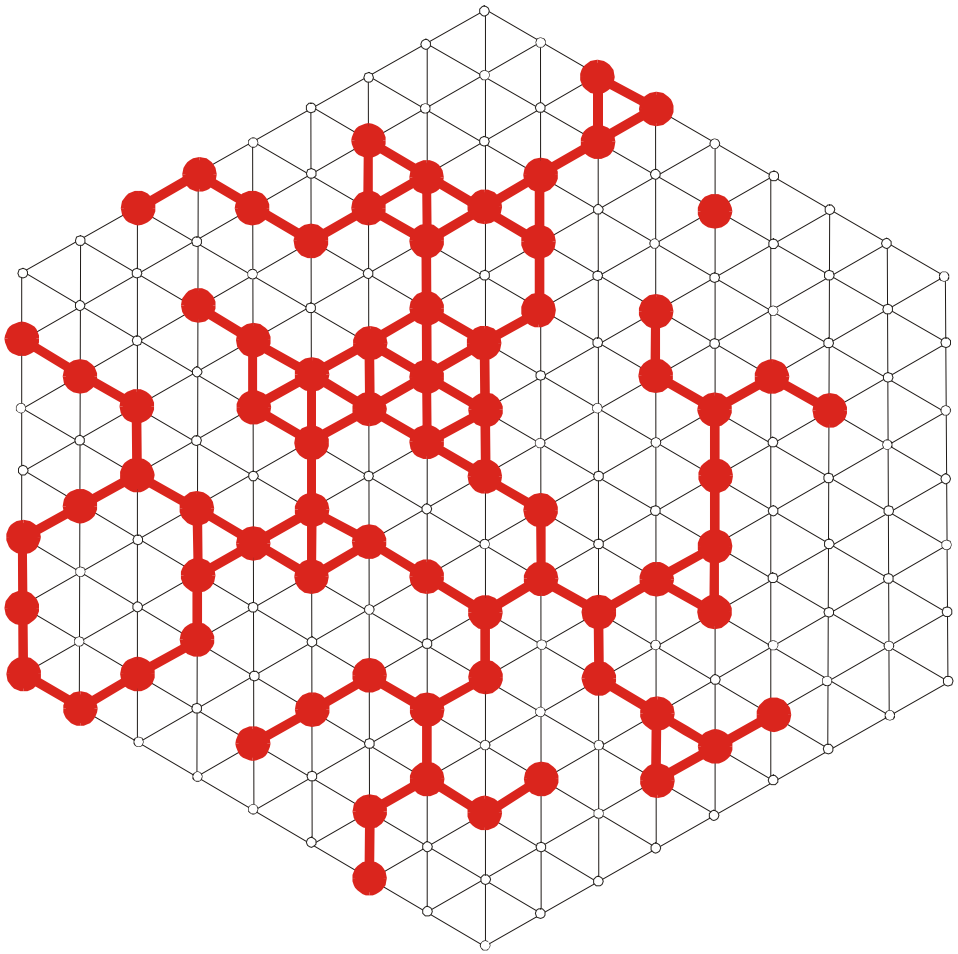
Sketch of sequence space



Random graph approach to neutral networks

Step 75

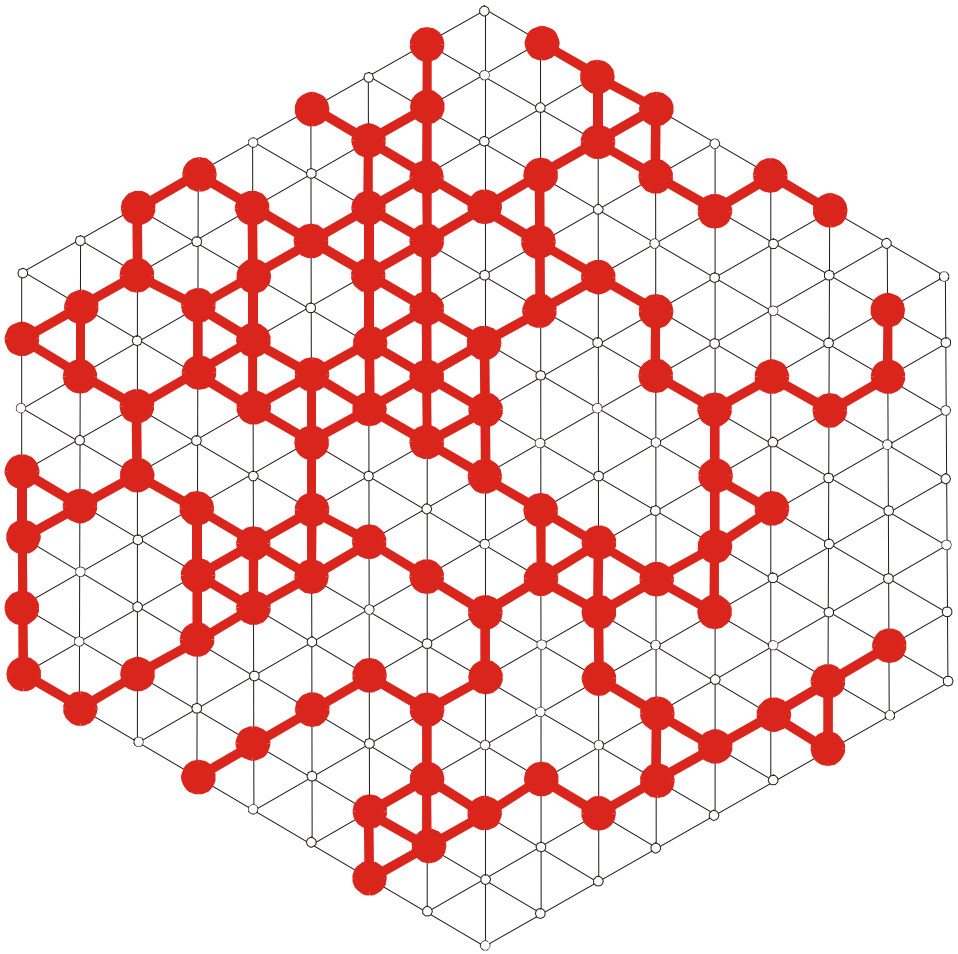
Sketch of sequence space



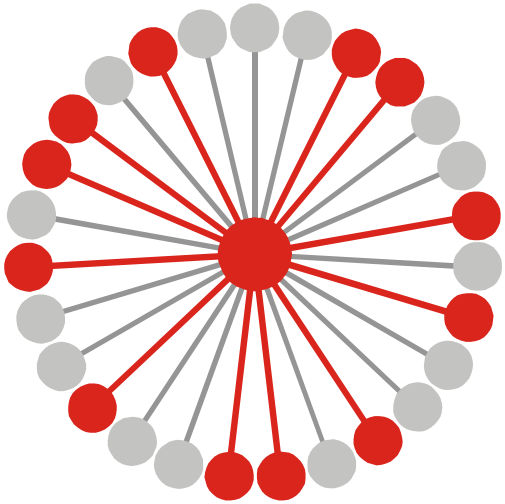
Random graph approach to neutral networks

Step 100

Sketch of sequence space



Random graph approach to neutral networks



$$G_k = y^{-1}(S_k)\ Y\ \{\ I_j \mid y(I_j) = S_k\ \}$$

$$\lambda_j = 12 / 27\ , \qquad \bar{\lambda}_k = \frac{\underset{j \in |G_k|}{\text{ø}}\ l_j(k)}{|G_k|}$$

Connectivity threshold: $\lambda_{cr} = 1 - \kappa^{-1/(\kappa-1)}$

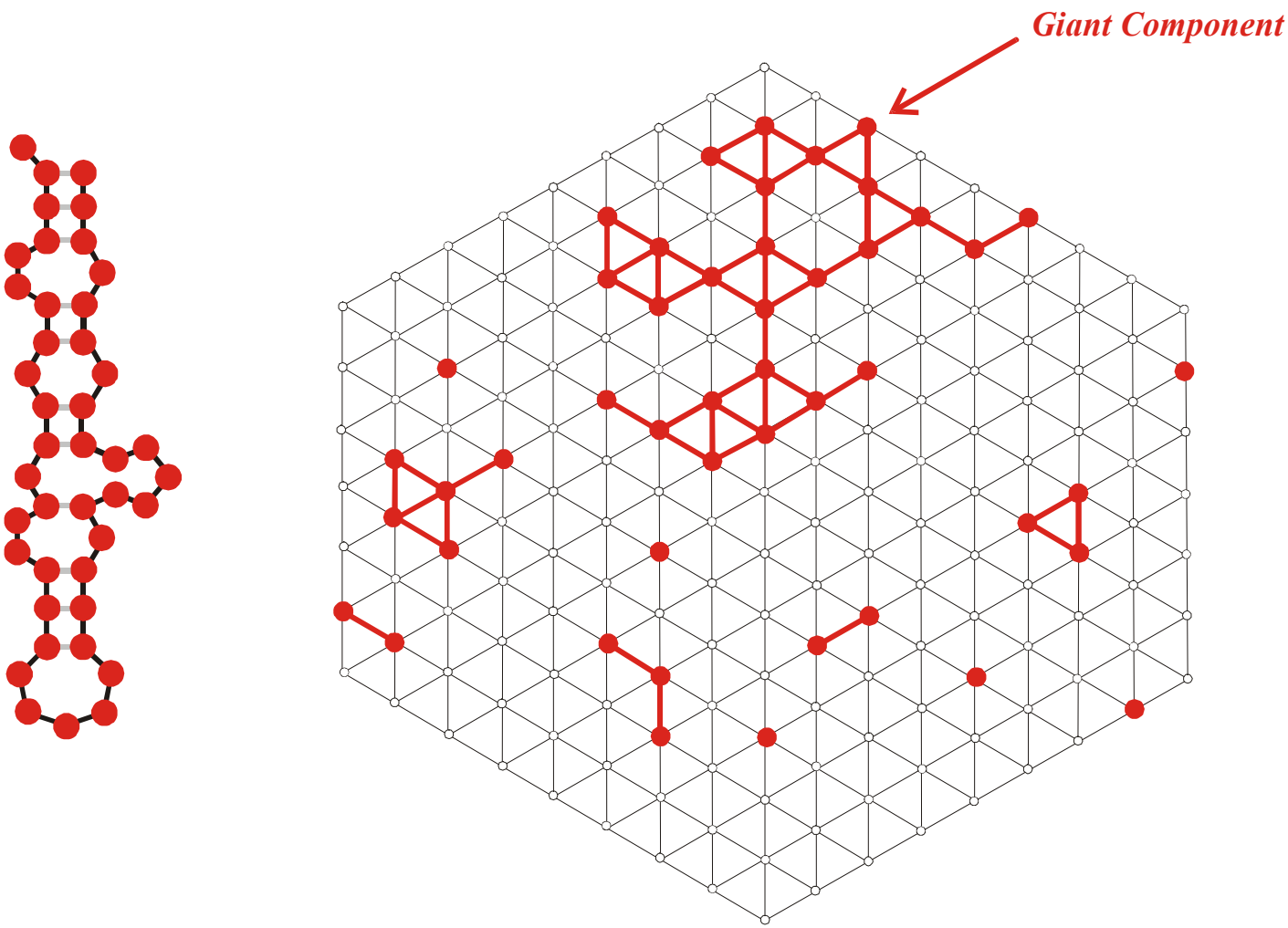
Alphabet size k : **AUGC** ñ k = 4

k	l_{cr}
2	0.5
3	0.4226
4	0.3700

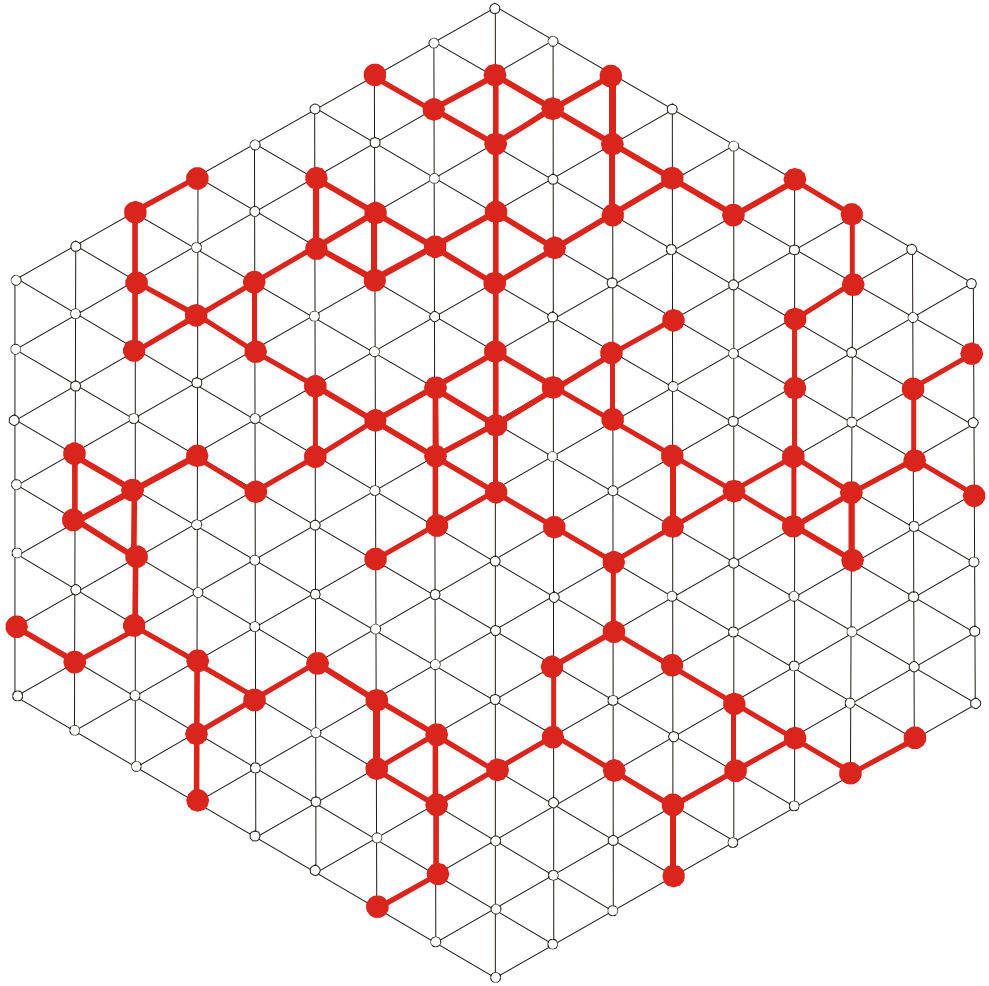
$\bar{\lambda}_k > \lambda_{cr}$ .... network G_k is connected

$\bar{\lambda}_k < \lambda_{cr}$ .... network G_k is **not** connected

Mean degree of neutrality and connectivity of neutral networks



A multi-component neutral network



A connected neutral network

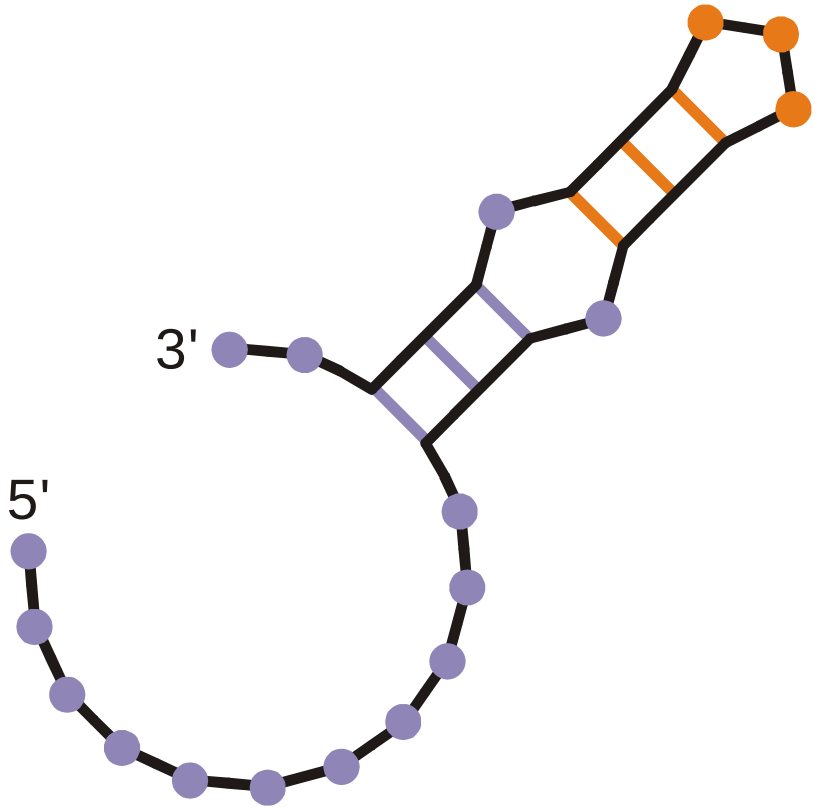
Suboptimal RNA Secondary Structures

Michael Zuker. *On finding all suboptimal foldings of an RNA molecule*. Science **244** (1989), 48-52

Stefan Wuchty, Walter Fontana, Ivo L. Hofacker, Peter Schuster. *Complete suboptimal folding of RNA and the stability of secondary structures.* Biopolymers **49** (1999), 145-165

Total number of structures including all suboptimal conformations, stable and unstable (with $DG_0>0$):

#conformations = **1 416 661**

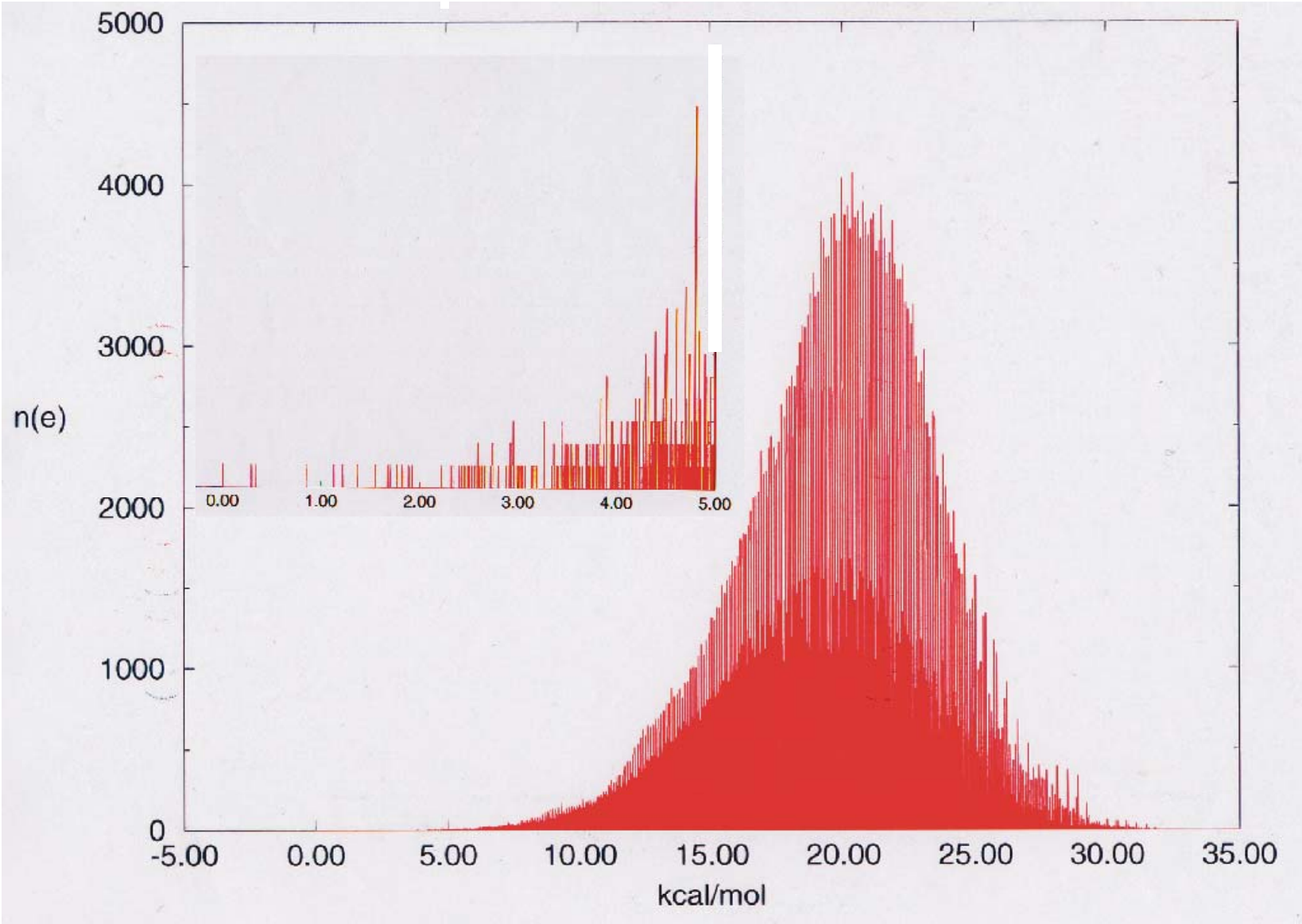


Minimum free energy structure

AAAGGGCACAGGGUGAUUUCAAUAAUUUUA

Sequence

Example of a small RNA molecule: n=30



Density of stares of suboptimal structures of the RNA molecule with the sequence:

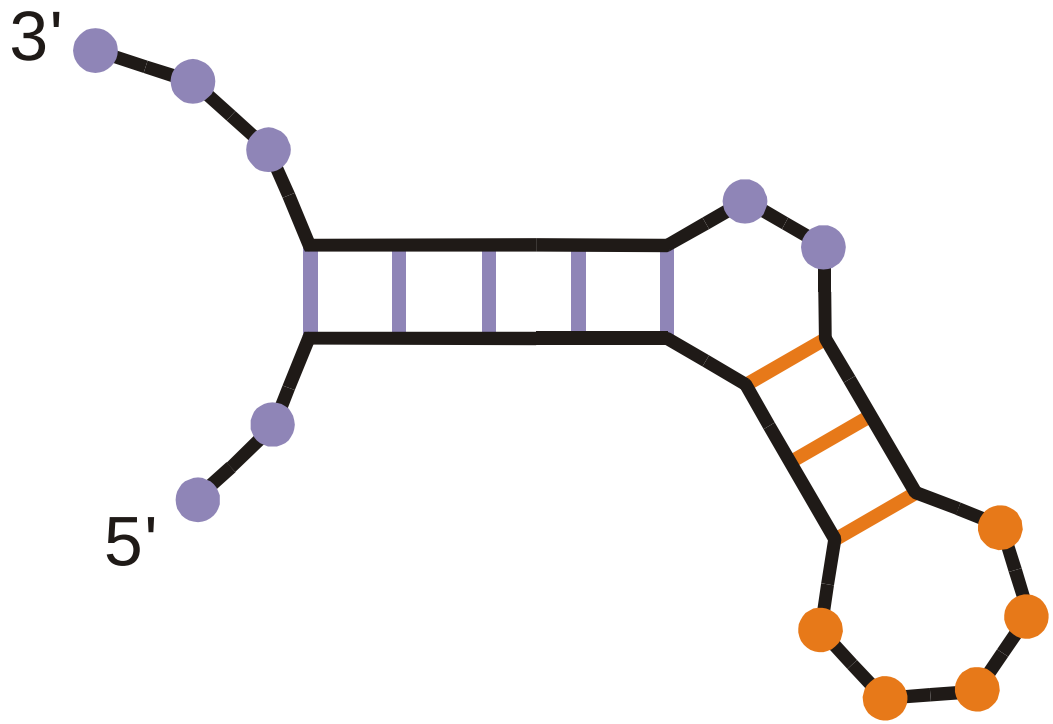
AAAGGGCACAGGGUGAUUUCAAUAAUUUUA

Partition Function of RNA Secondary Structures

John S. McCaskill. *The equilibrium function and base pair binding probabilities for RNA secondary structure*. Biopolymers **29** (1990), 1105-1119

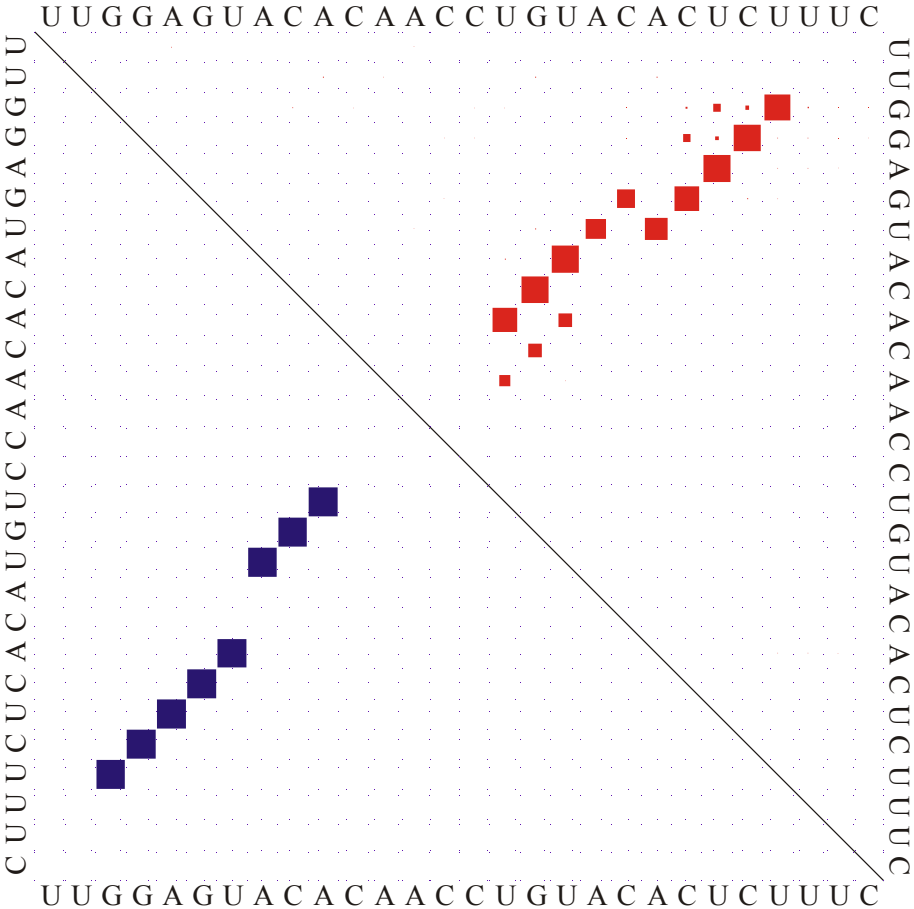
Ivo L. Hofacker, Walter Fontana, Peter F. Stadler, L. Sebastian Bonhoeffer, Manfred Tacker, Peter Schuster. *Fast folding and comparison of RNA secondary structures.* Monatshefte für Chemie **125** (1994), 167-188

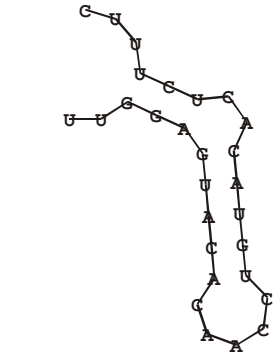
Example of a small RNA molecule with two low-lying suboptimal conformations which contribute substantially to the partition function



UUGGAGUACACAACCUGUACACUCUUUC

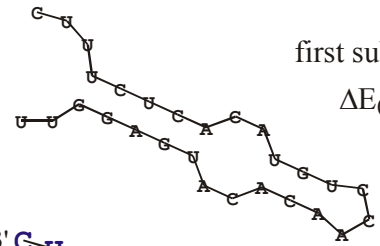
Example of a small RNA molecule: n=28





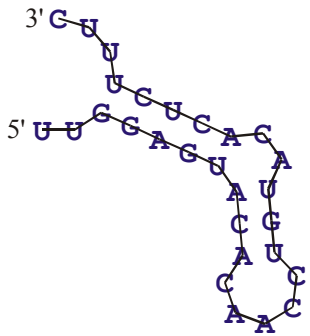
second suboptimal configuration

$\Delta E_{0\rightarrow 2}$ = 0.55 kcal / mole



first suboptimal configuration

$\Delta E_{0\rightarrow 1}$ = 0.50 kcal / mole



minimum free energy configuration

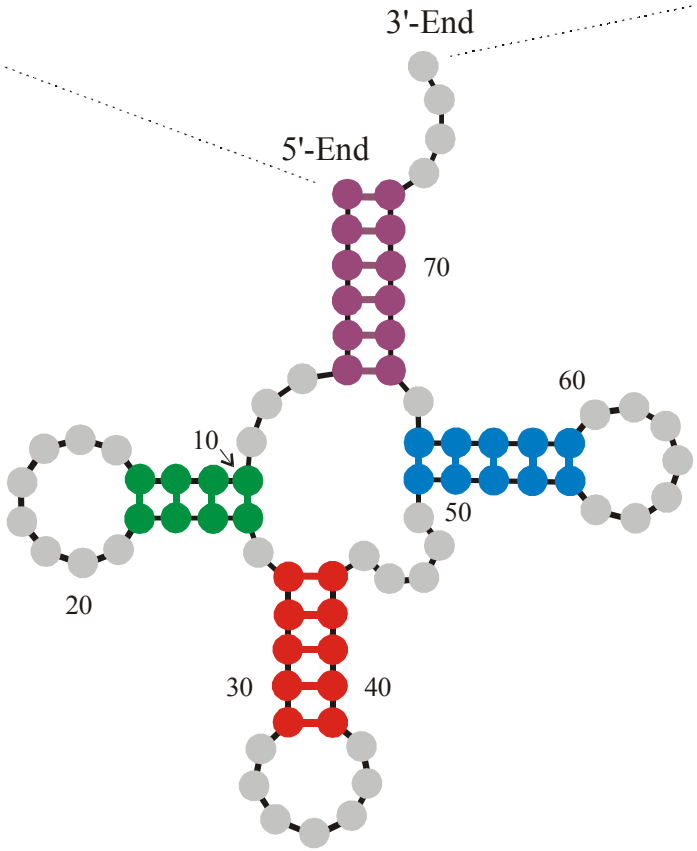
DG_0 = - 5.39 kcal / mole

„Dot plot“ of the minimum free energy structure (**lower triangle**) and the partition function (**upper triangle**) of a small RNA molecule (n=28) with low energy suboptimal configurations

Sequence

5'-End GCGGAUUUAGCUCAGDDGGGAGAGCMCCAGACUGAAYAUCUGGAGMUCCUGUGTPCGAUCCACAGAAUUCGCACCA 3'-End

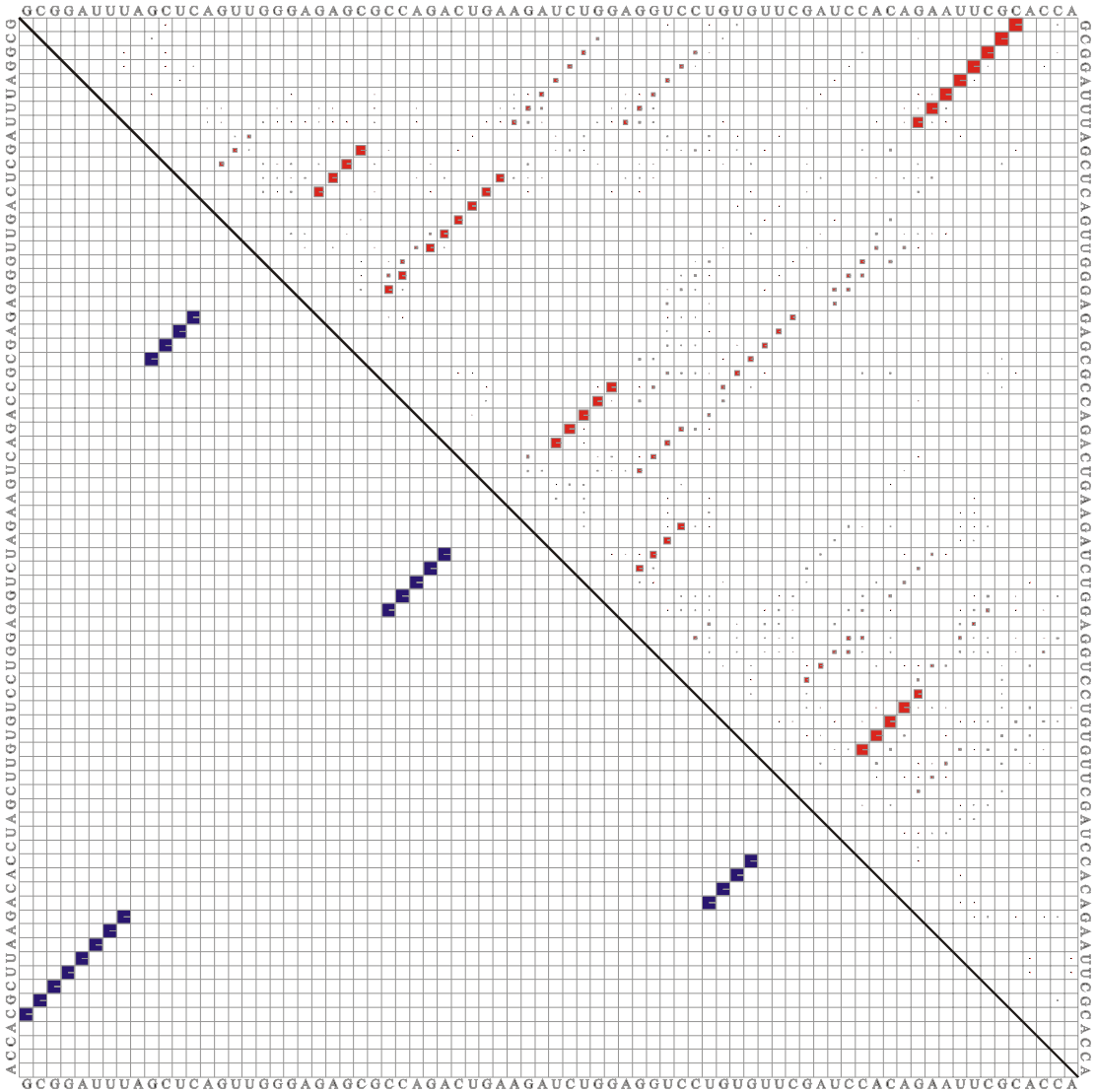
Secondary Structure

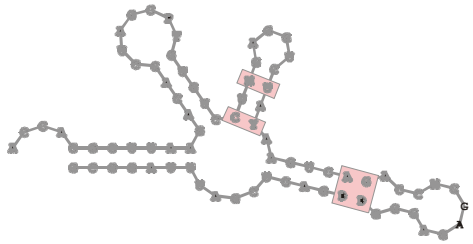


Symbolic Notation

5'-End ((((((···((((·····))))·((((·····))))).....((((·····))))).))))···· 3'-End

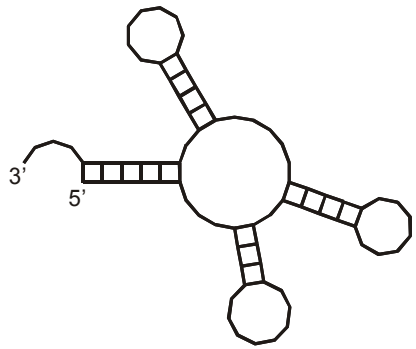
Phenylalanyl-tRNA as an example for the computation of the partition function





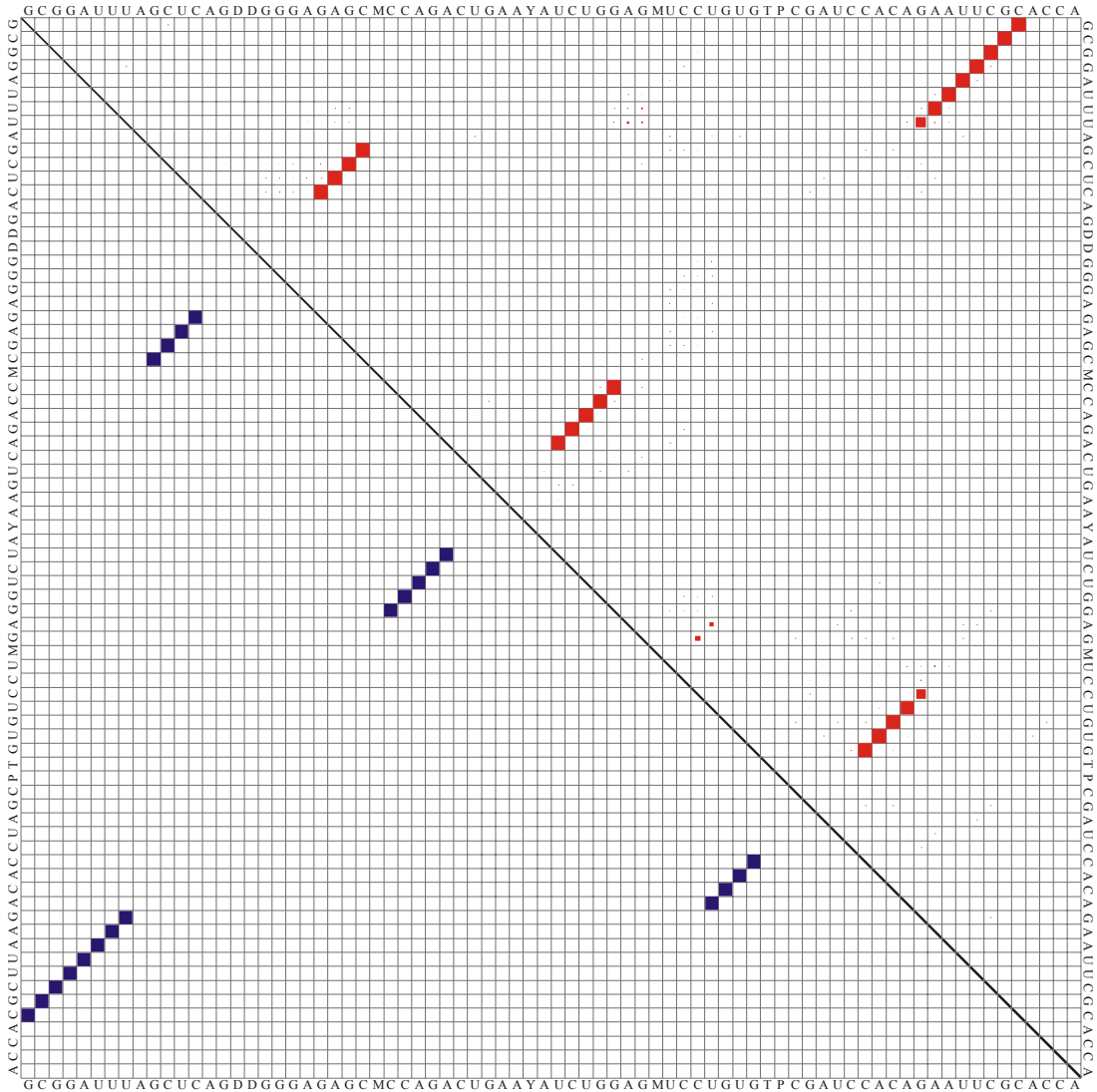
first suboptimal configuration

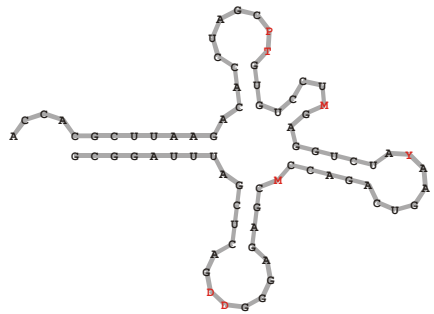
$\Delta E_{0\rightarrow 1}$ = 0.43 kcal / mole



tRNA^{phe}

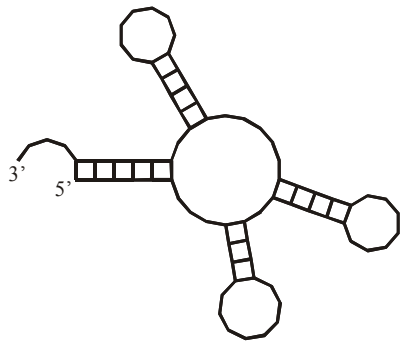
without modified bases





first suboptimal configuration

$\Delta E_{0\rightarrow 1}$ = 0.94 kcal / mole



tRNA^{phe}

with modified bases

Kinetic Folding of RNA Secondary Structures

Christoph Flamm, Walter Fontana, Ivo L. Hofacker, Peter Schuster. *RNA folding kinetics at elementary step resolution.* RNA **6**:325-338, 2000

Christoph Flamm, Ivo L. Hofacker, Sebastian Maurer-Stroh, Peter F. Stadler, Martin Zehl. *Design of multistable RNA molecules.* RNA **7**:325-338, 2001

The Folding Algorithm

A sequence I specifies an energy ordered set of compatible structures S(I):

$$S(I) = \{S_0, S_1, \ldots, S_m, \mathbf{O}\}$$

A trajectory $T_k(I)$ is a time ordered series of structures in S(I). A folding trajectory is defined by starting with the open chain **O** and ending with the global minimum free energy structure S_0 or a metastable structure S_k which represents a local energy minimum:

$$T_0(I) = \{\mathbf{O}, S(1), \ldots, S(t\text{-}1), S(t), S(t+1), \ldots, \mathbf{S_0}\}$$
$$T_k(I) = \{\mathbf{O}, S(1), \ldots, S(t\text{-}1), S(t), S(t+1), \ldots, \mathbf{S_k}\}$$

Transition probabilities $P_{ij}(t) = \mathrm{Prob}\{S_i \rightarrow S_j\}$ are defined by

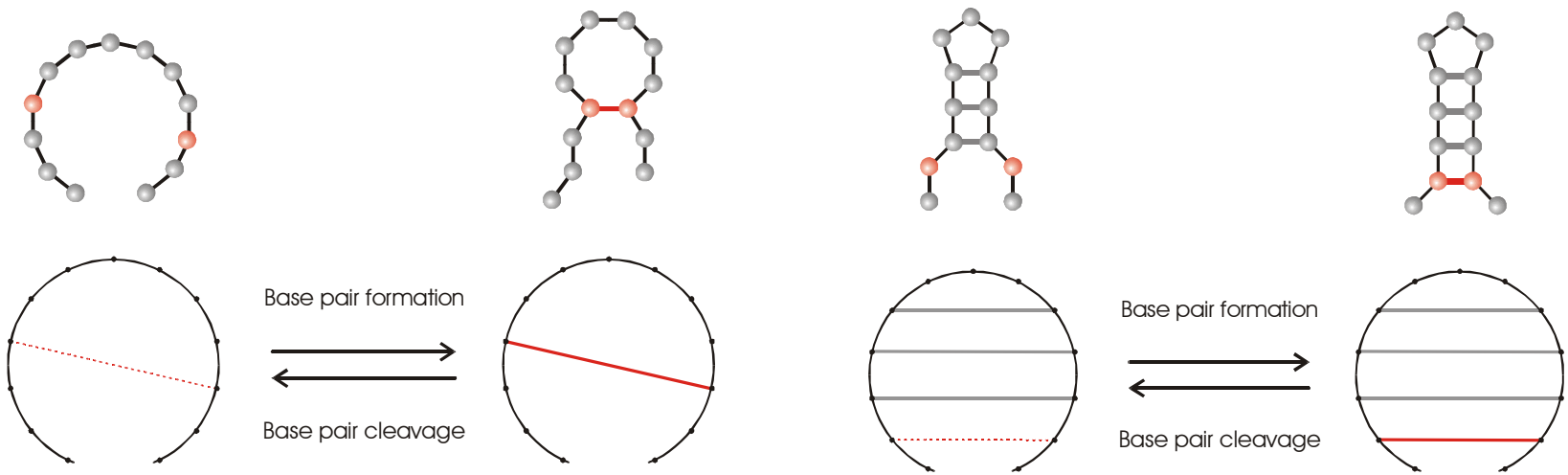
$$P_{ij}(t) = P_i(t)\, k_{ij} = P_i(t) \exp(-\Delta G_{ij}/2RT) / \Sigma_i$$

$$P_{ji}(t) = P_j(t)\, k_{ji} = P_j(t) \exp(-\Delta G_{ji}/2RT) / \Sigma_j$$

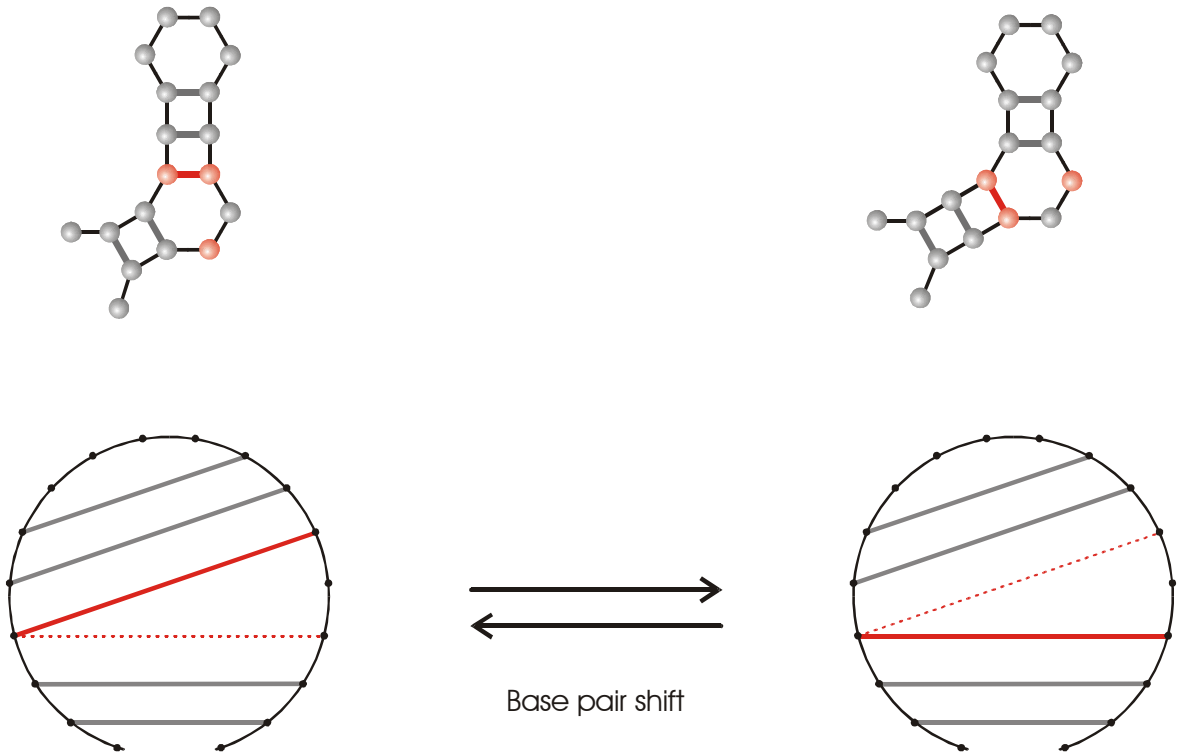
$$\Sigma_k = \sum_{k=1, k\neq i}^{m+2} \exp(-\Delta G_{ki}/2RT)$$

The symmetric rule for transition rate parameters is due to Kawasaki (K. Kawasaki, *Diffusion constants near the critical point for time depen-dent Ising models*. Phys.Rev. **145**:224-230, 1966).

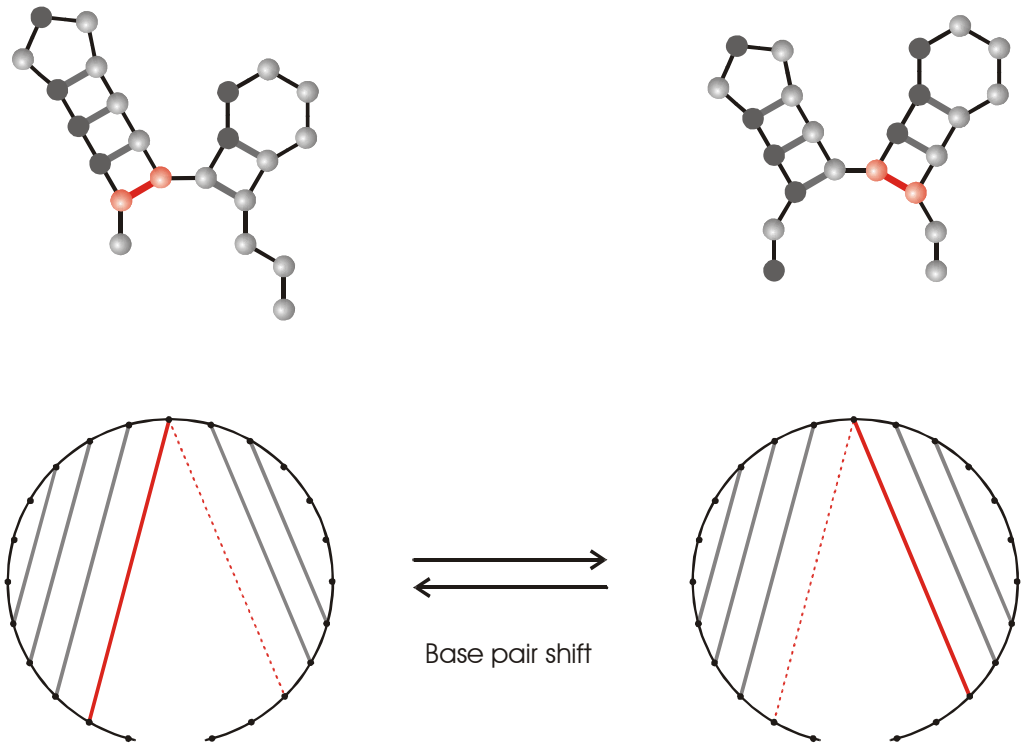
Formulation of kinetic RNA folding as a stochastic process



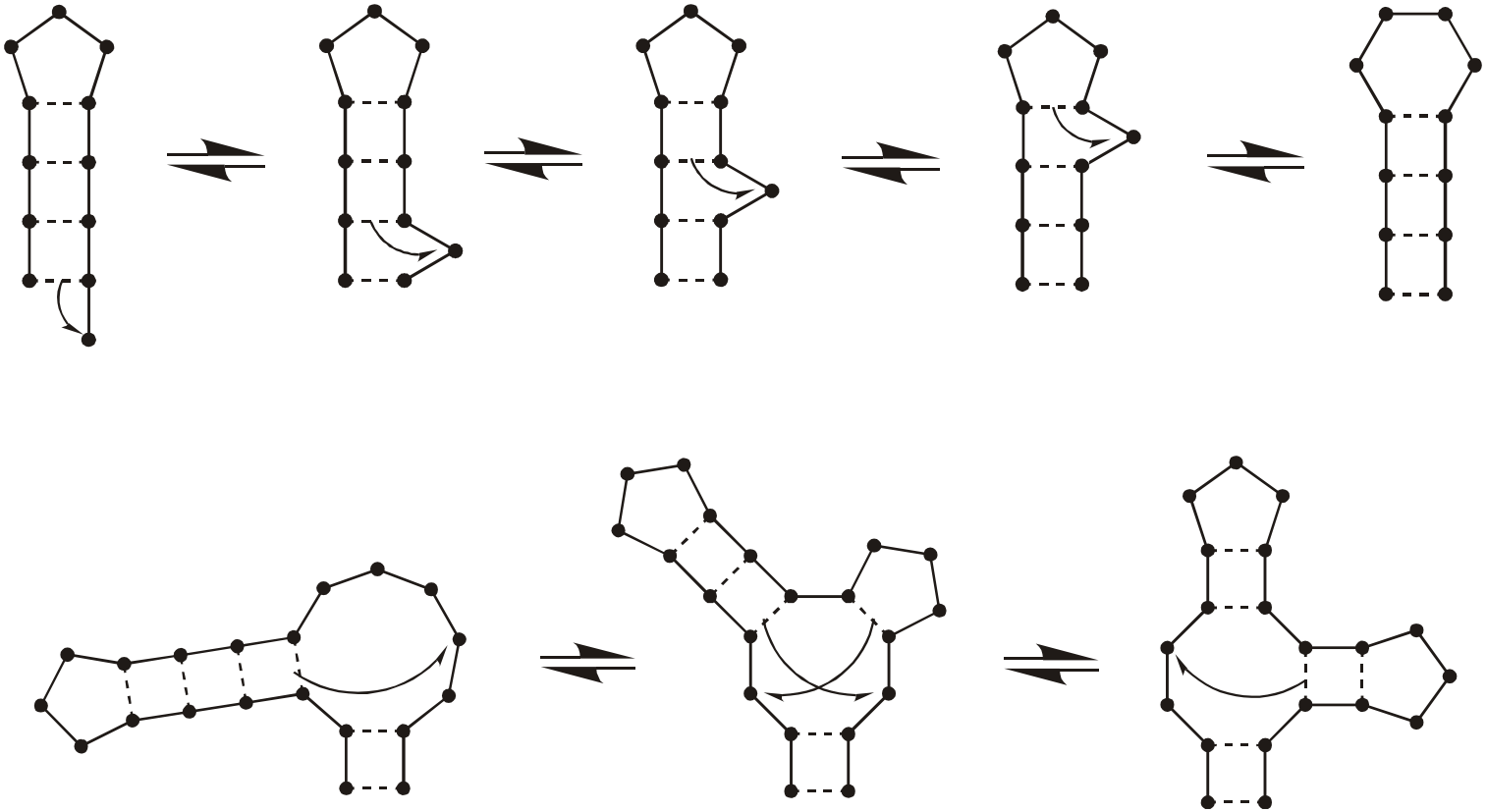
Base pair formation and base pair cleavage moves for nucleation and elongation of stacks



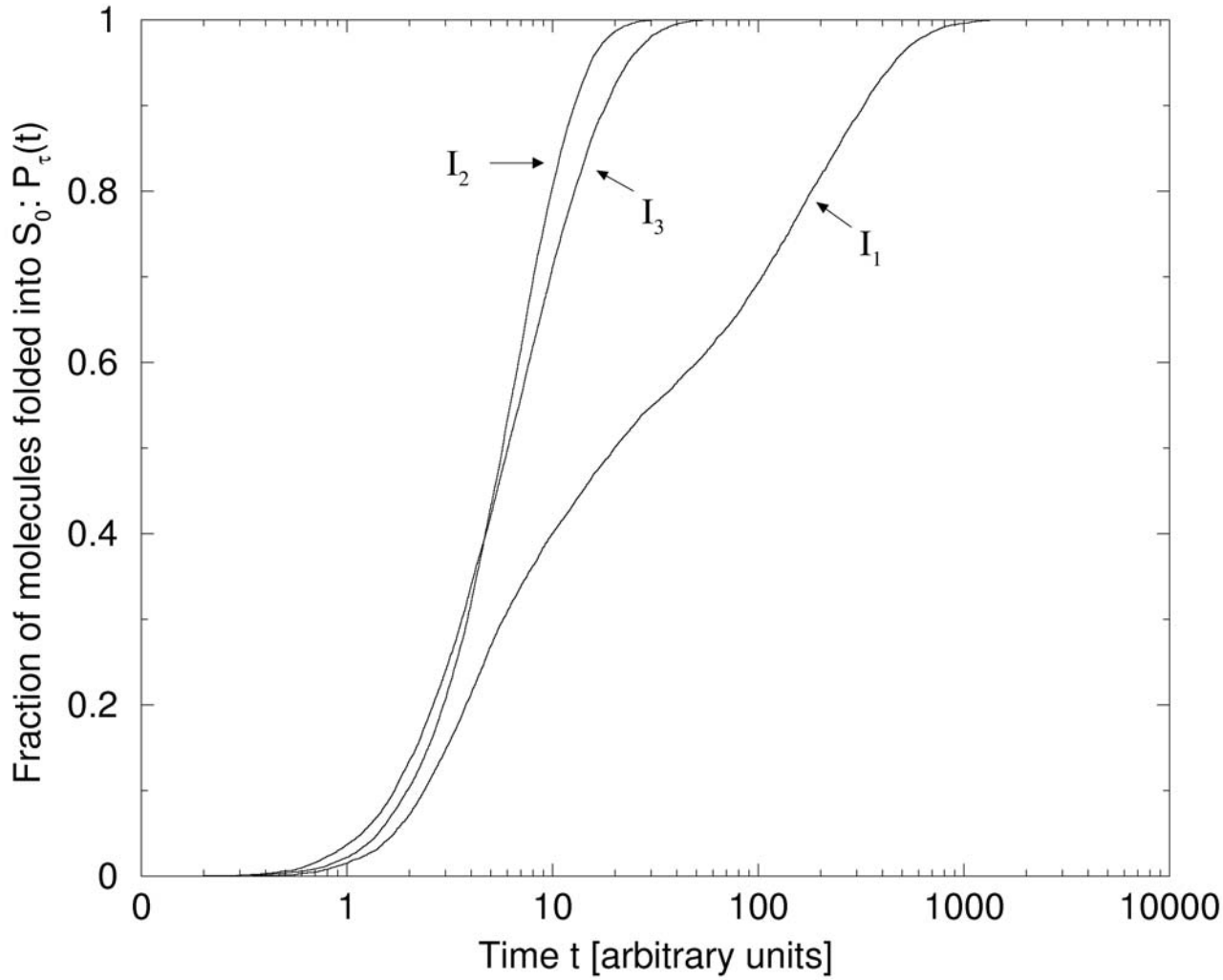
Base pair shift move of class 1: Shift inside internal loops or bulges



Base pair shift move of class 2: Shift involving free ends

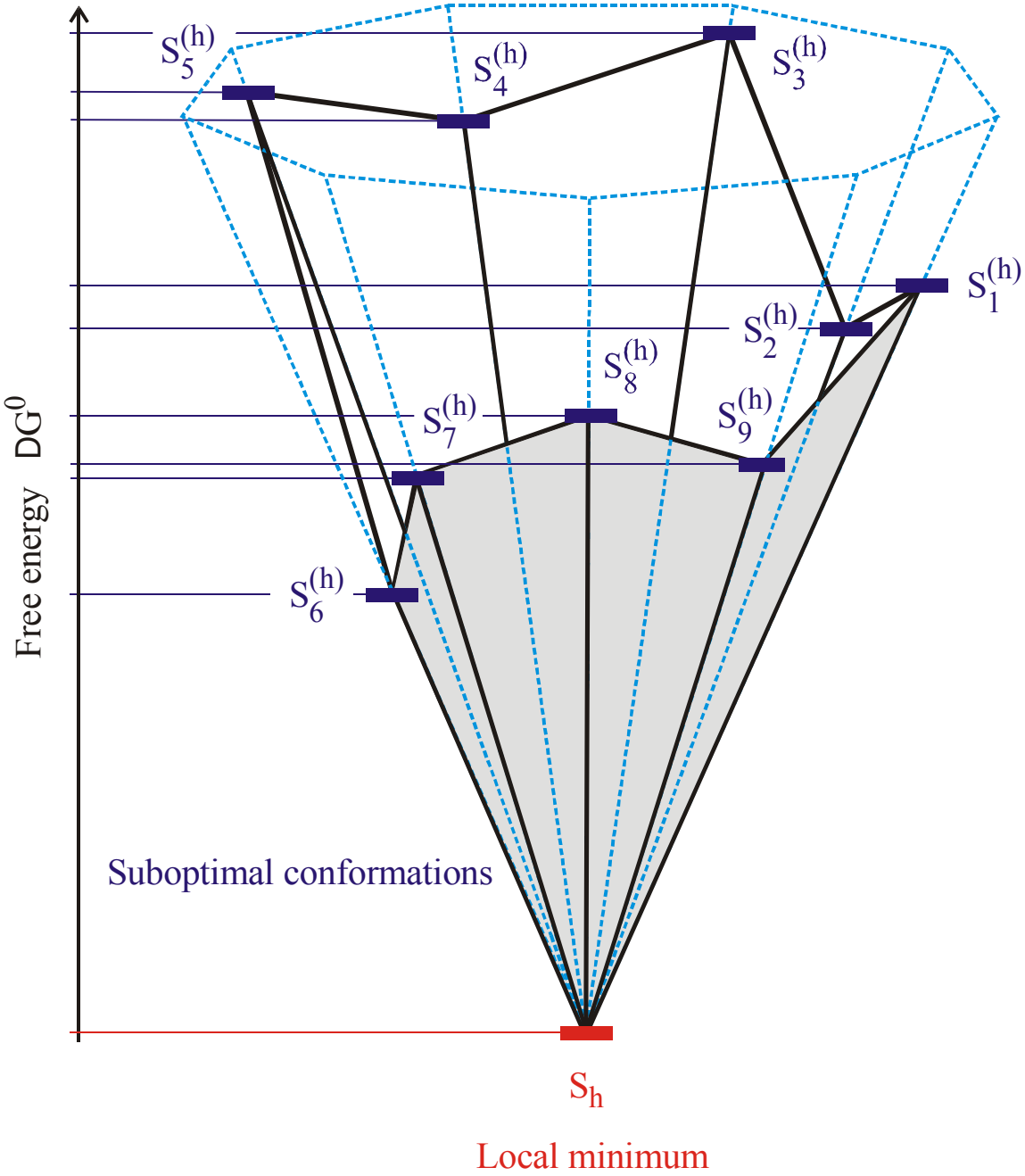


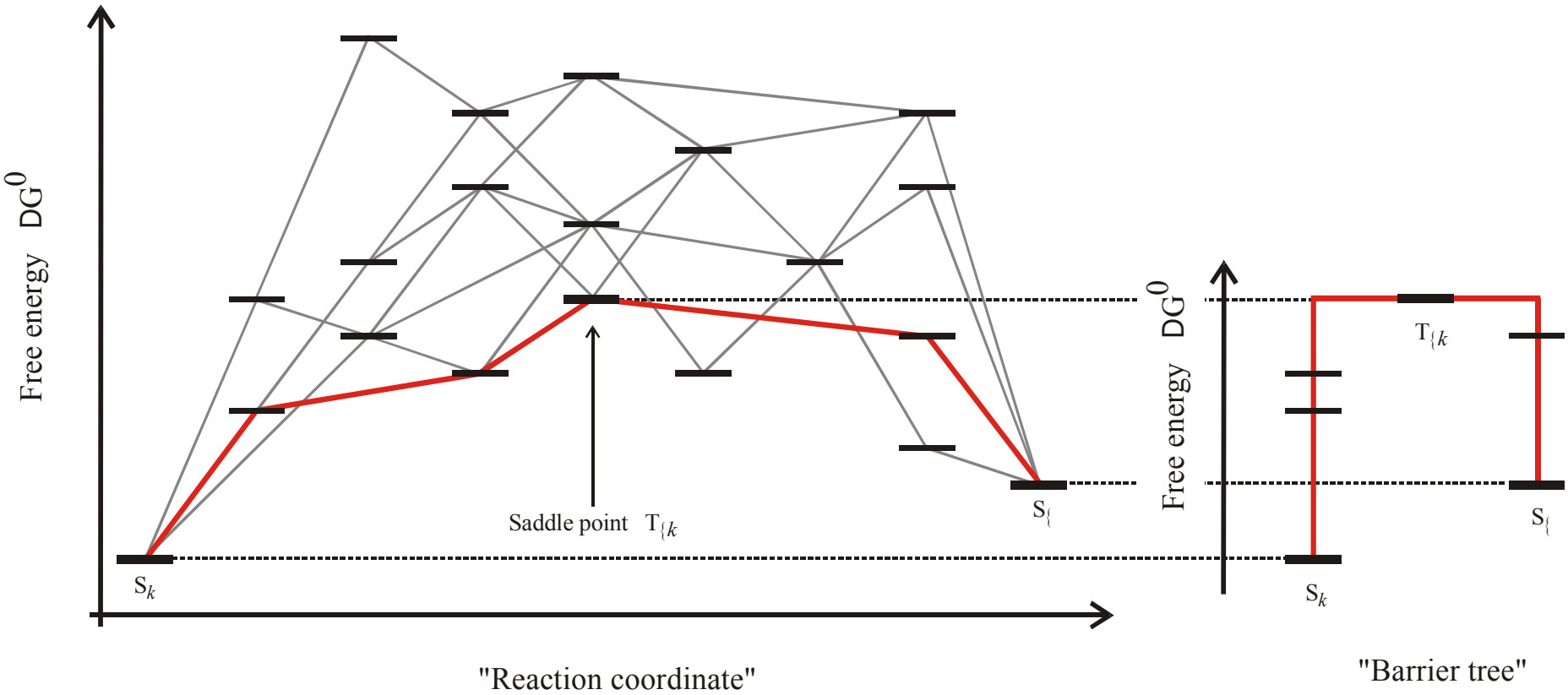
Examples of rearrangements through consecutive shift moves

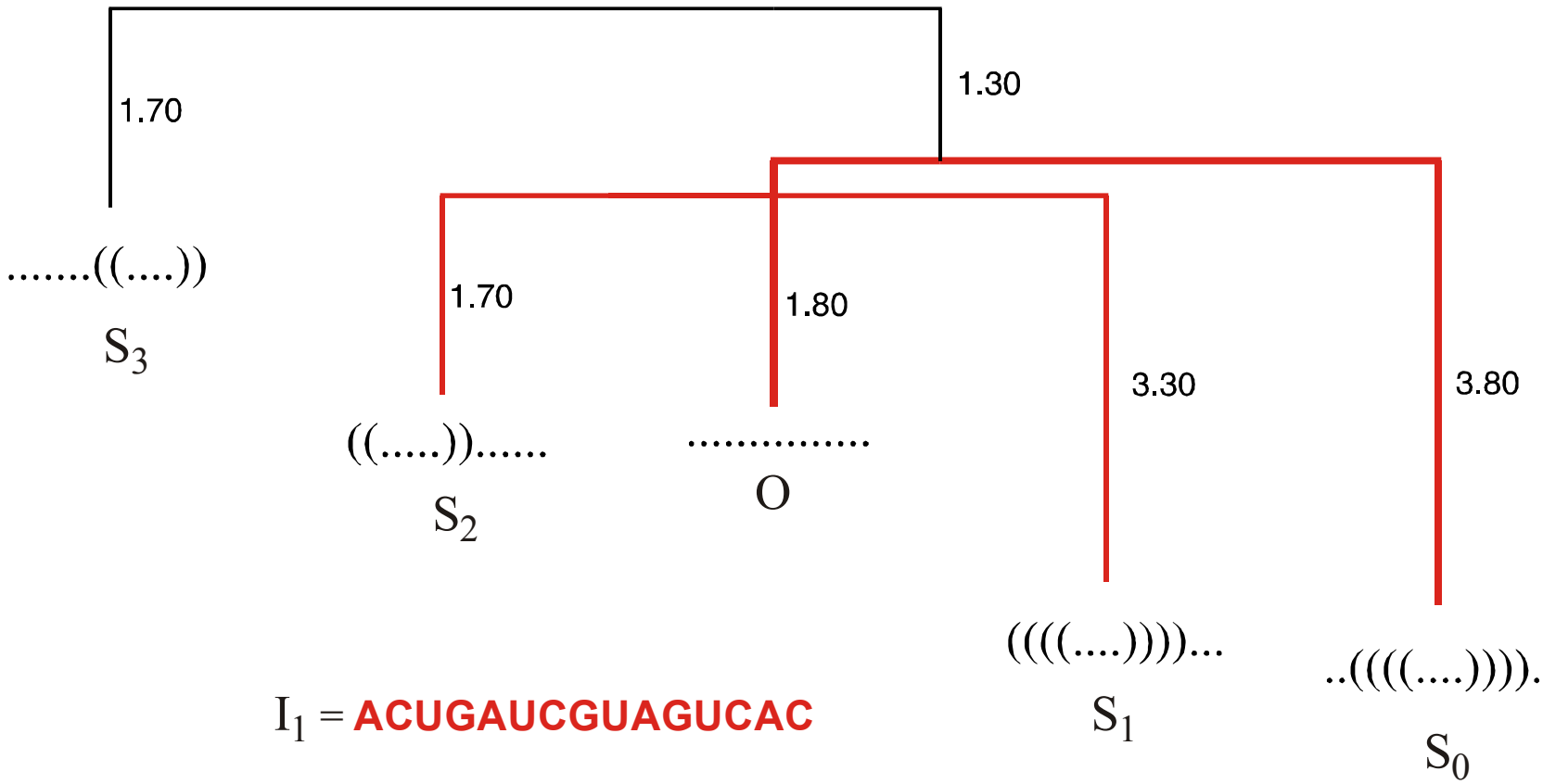


Mean folding curves for three small RNA molecules with different folding behavior

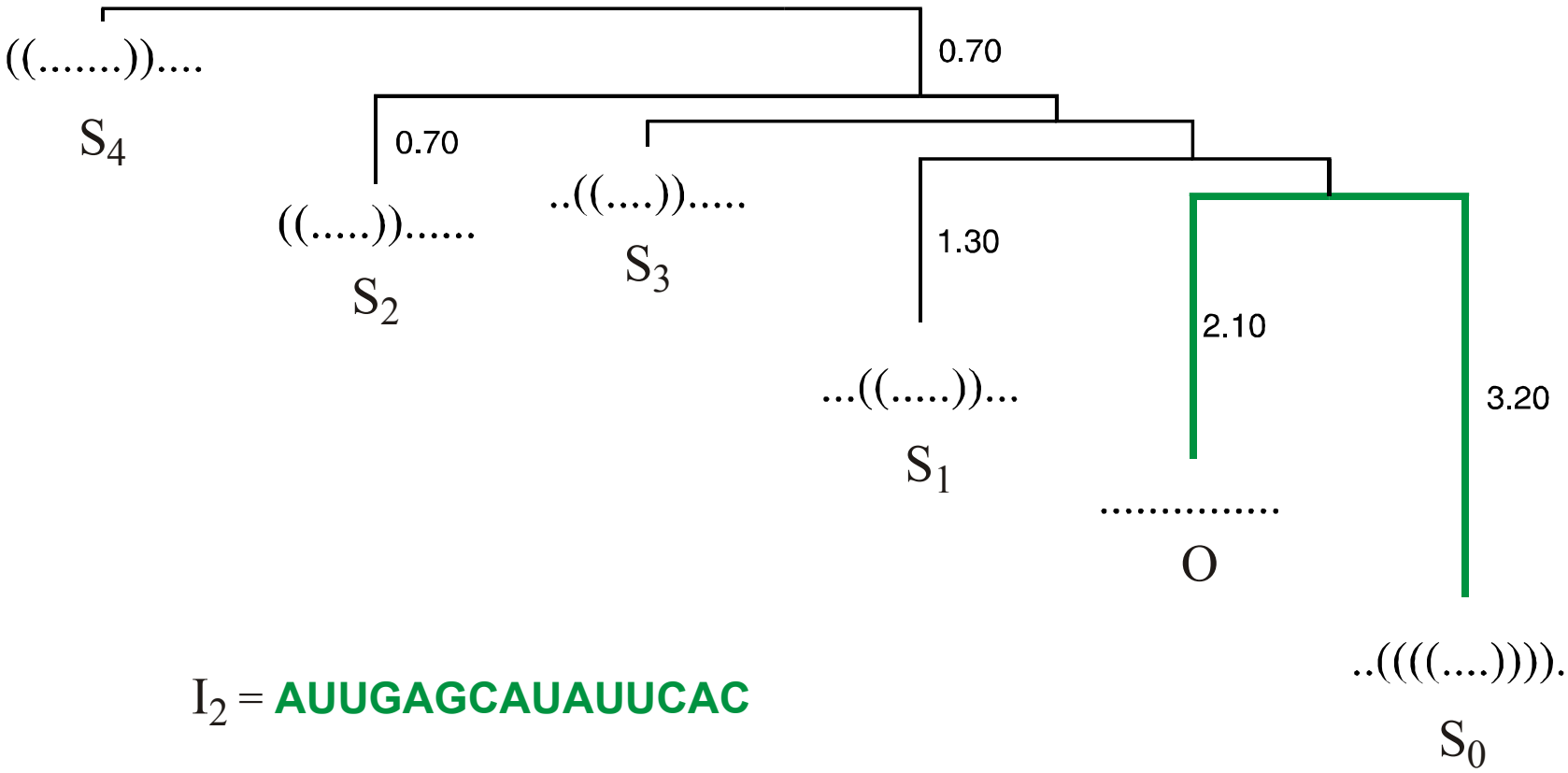
Search for local minima in conformation space




Free energy DG0
Saddle point T{k
S{
Sk
"Reaction coordinate"
Free energy DG0
T{k
S{
Sk
"Barrier tree"

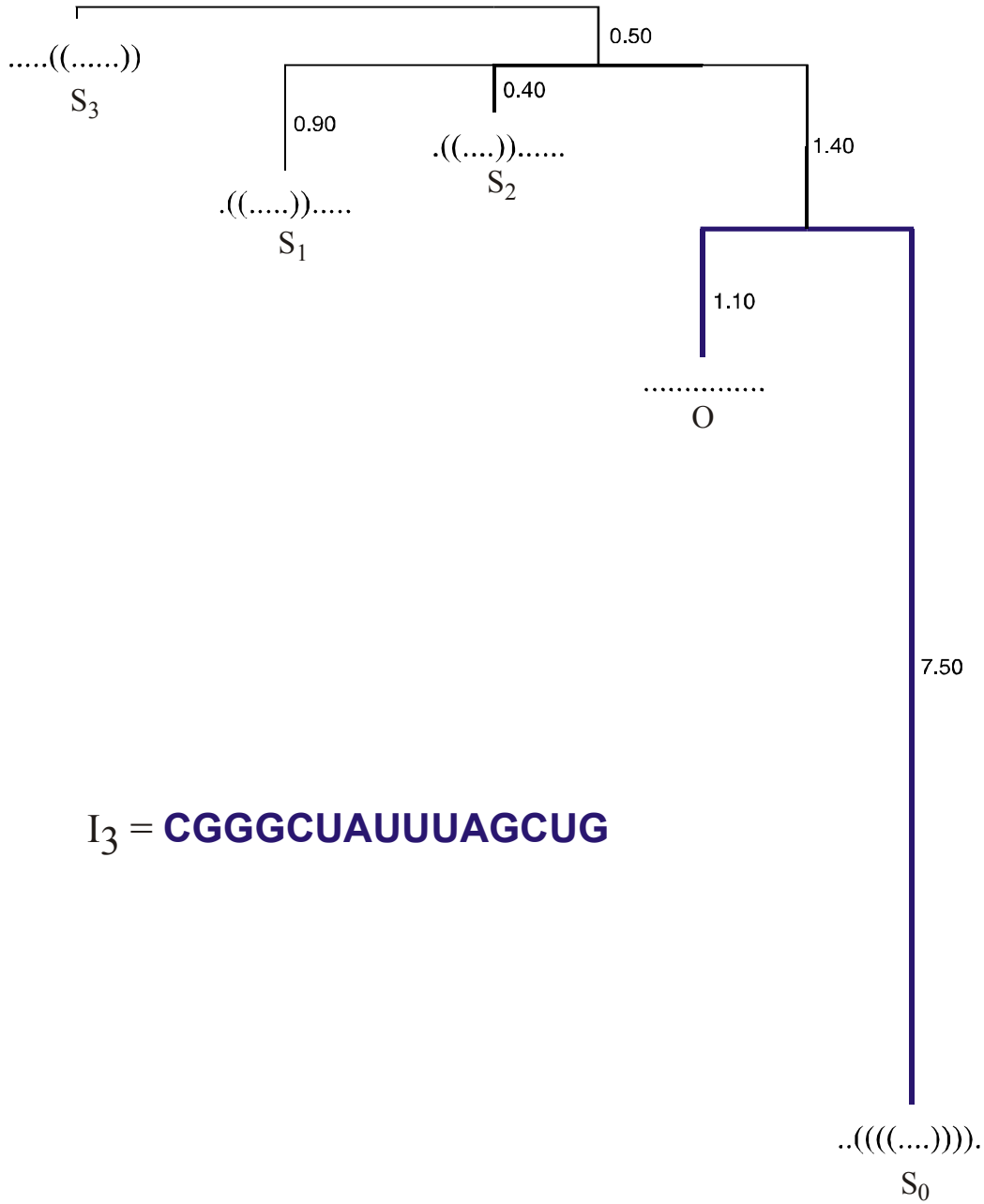


Example of an unefficiently folding small RNA molecule with n = 15



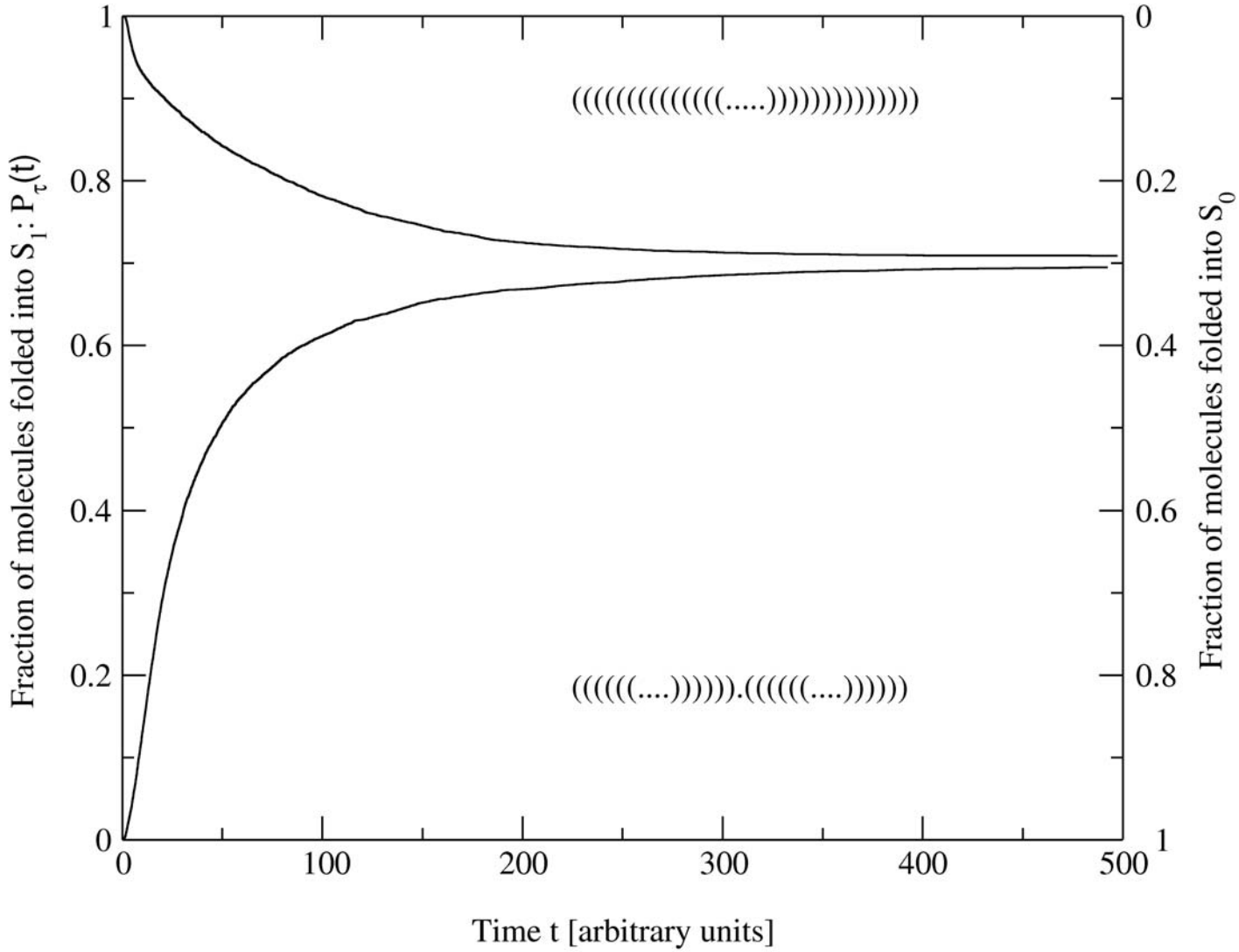
I_2 = **AUUGAGCAUAUUCAC**

Example of an easily folding small RNA molecule with n = 15



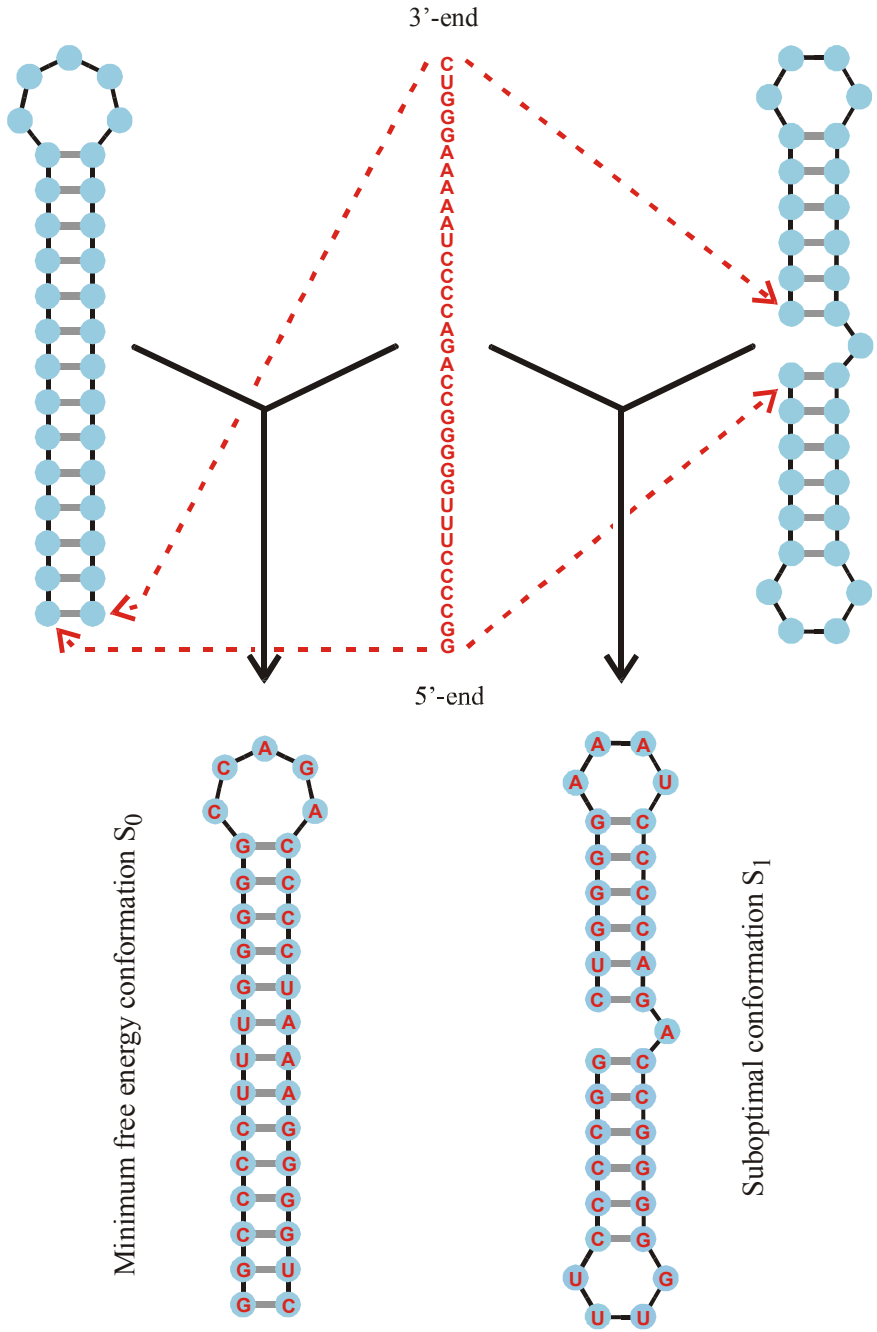
I_3 = **CGGGCUAUUUAGCUG**

Example of an easily folding and especially stable small RNA molecule with n = 15

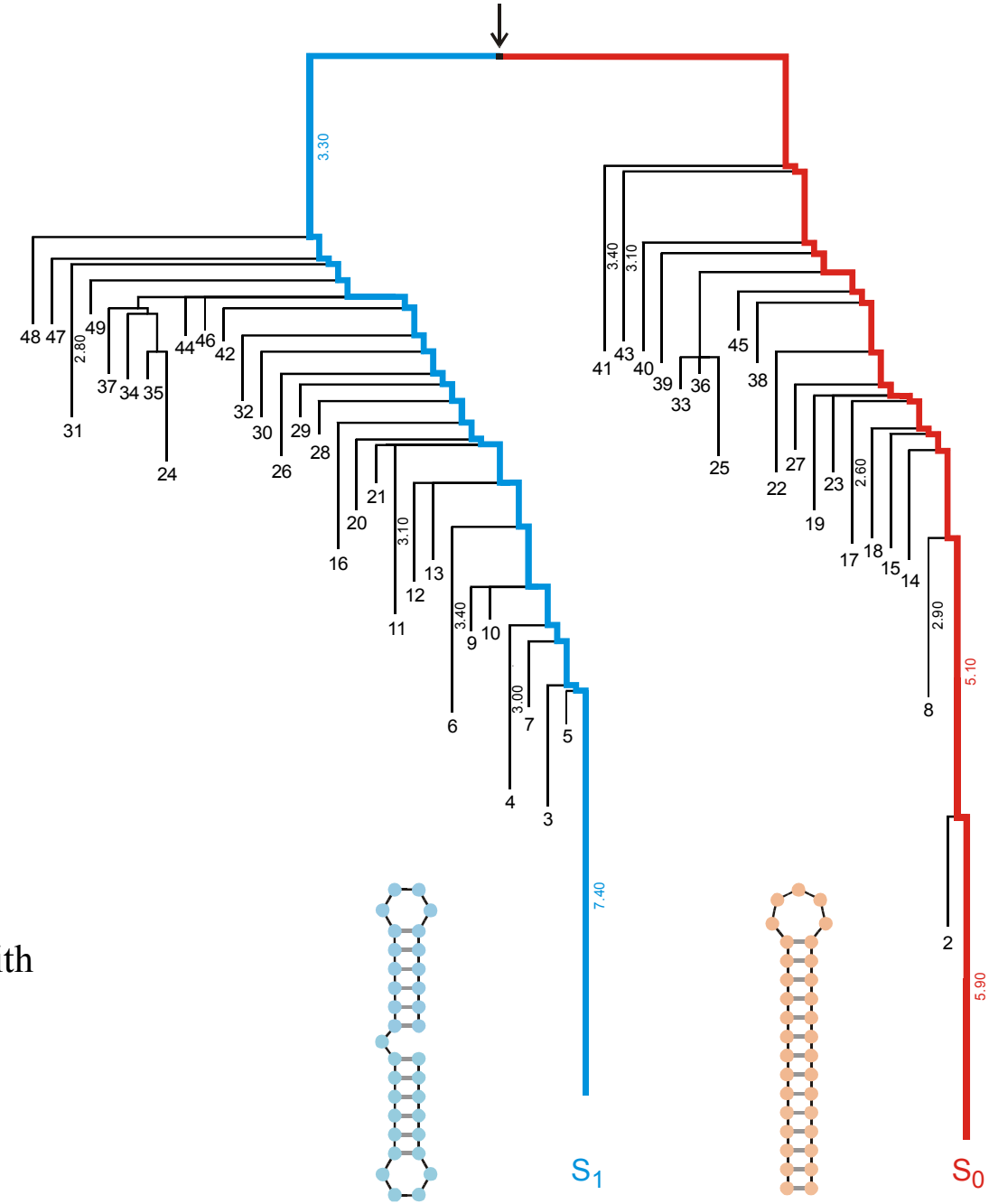


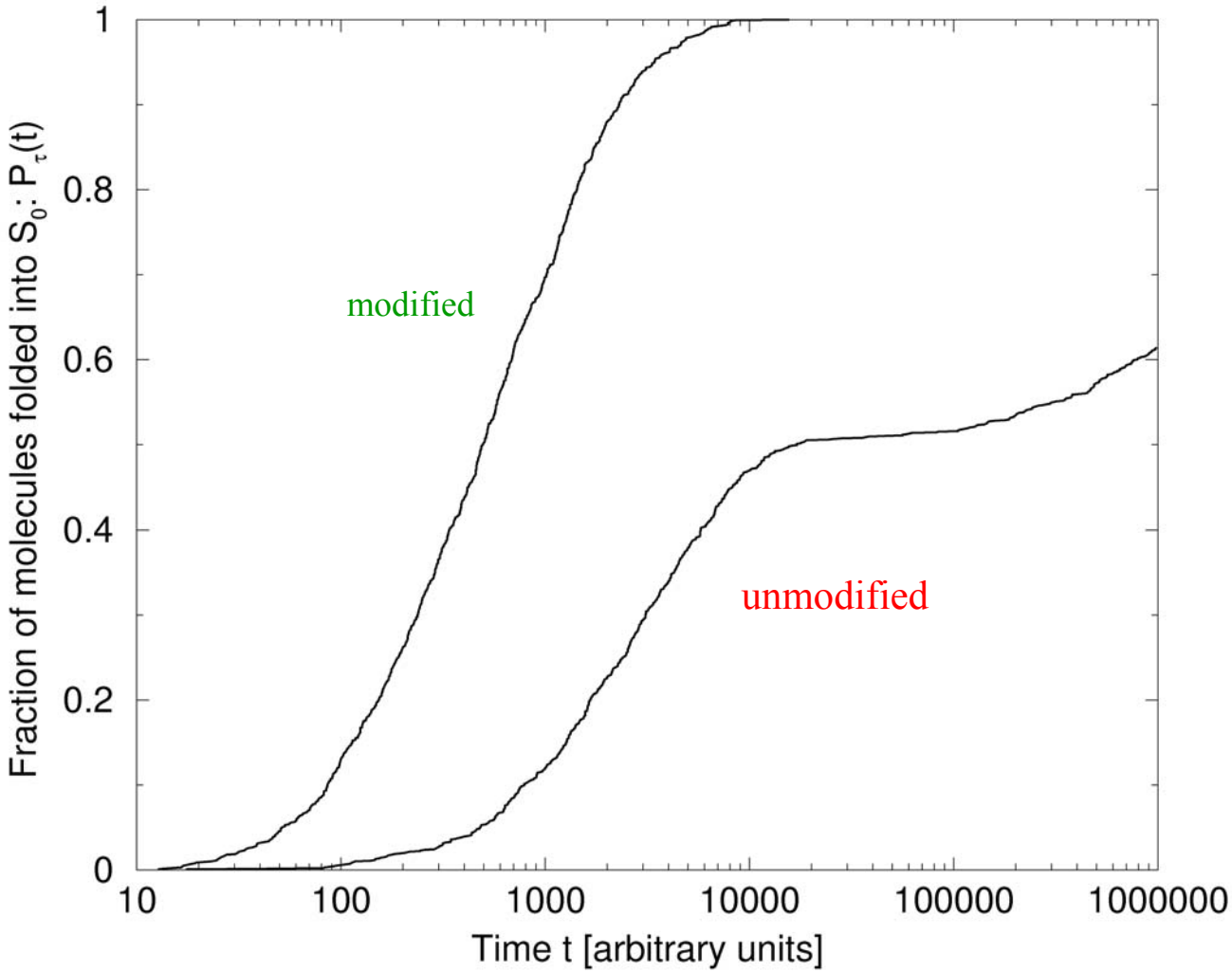
Folding dynamics of the sequence **GGCCCCUUUGGGGGCCAGACCCCUAAAAAGGGUC**

One sequence is compatible with two structures

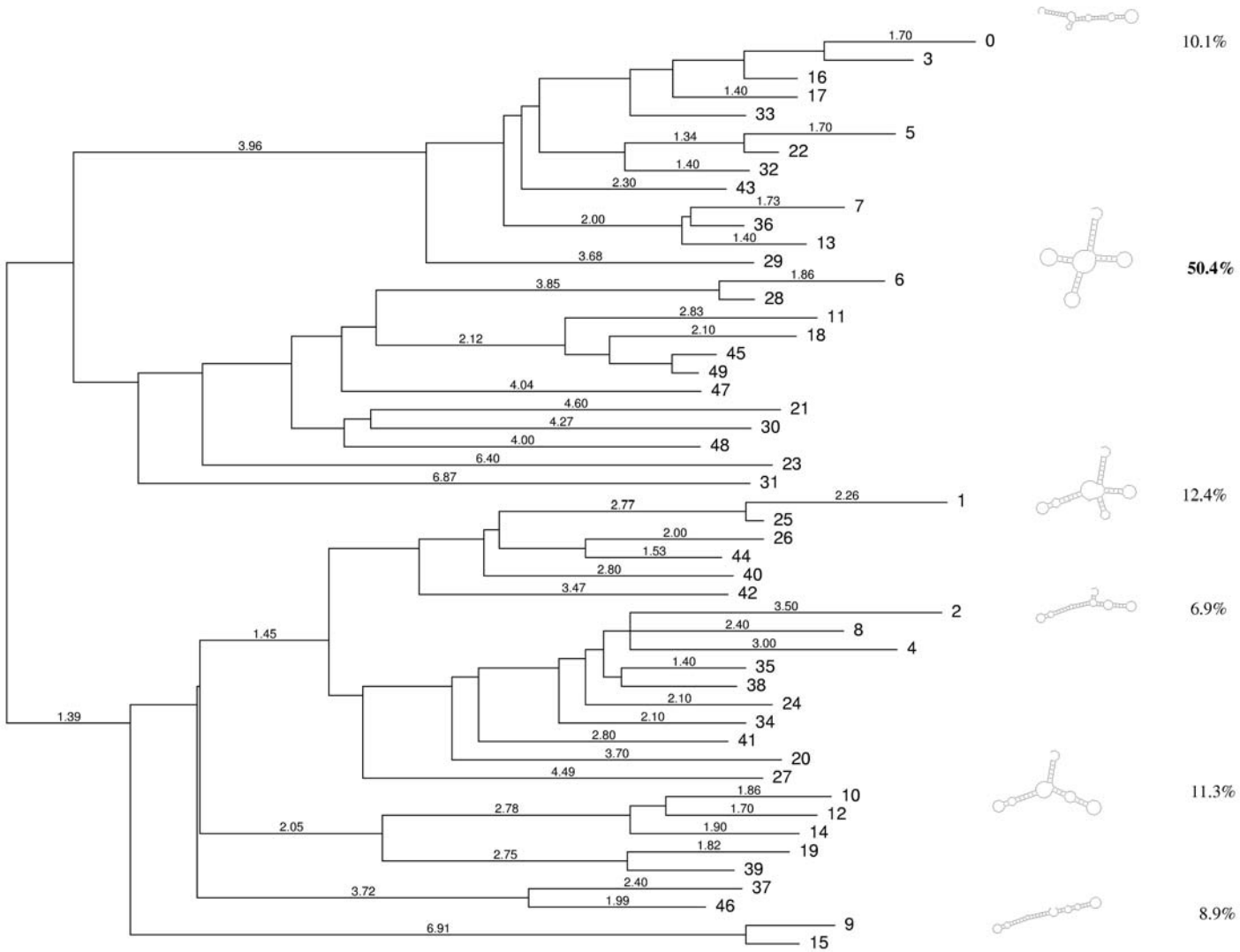


Barrier tree of a sequence with two conformations





Folding dynamics of tRNA^{phe} with and without modified nucelotides



Barrier tree of tRNA^{phe} without modified nucelotides

Coworkers

Walter Fontana, Santa Fe Institute, NM

Christian Reidys, Christian Forst, Los Alamos National Laboratory, NM

Peter Stadler, Universität Leipzig, GE

Ivo L.Hofacker, Christoph Flamm, Universität Wien, AT

Bärbel Stadler, Andreas Wernitznig, Universität Wien, AT
Michael Kospach, Ulrike Langhammer, Ulrike Mückstein, Stefanie Widder
Jan Cupal, Kurt Grünberger, Andreas Svrček-Seiler, Stefan Wuchty

Ulrike Göbel, Institut für Molekulare Biotechnologie, Jena, GE
Walter Grüner, Stefan Kopp, Jaqueline Weber