More than 40 Years Research on (Bio)Polymers

DNA the ‚star' among the biomolecules and RNA the ‚magic molecule'

Peter Schuster

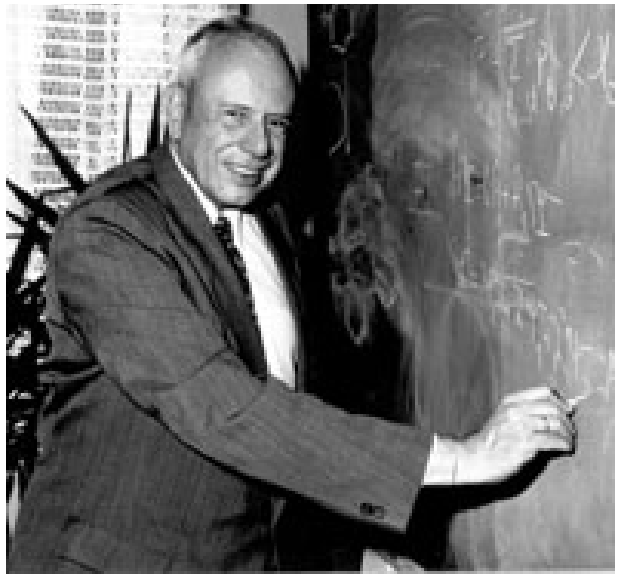
Institut für Theoretische Chemie, Universität Wien, Austria
and
The Santa Fe Institute, Santa Fe, New Mexico, USA

Central European Symposium for Theoretical Chemistry 2009

Dobogókő, 25.– 28.09.2009

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks



Born June 02, 1929 in Budapest

Happy birthday Janos !

PHYSICAL REVIEW VOLUME 155, NUMBER 3 15 MARCH 1967

Self-Consistent-Field Tight-Binding Treatment of Polymers. I. Infinite Three-Dimensional Case

G. DEL RE AND J. LADIK*
Gruppo Chimica Teorica del CNR, Via Celso 7, Rome and Istituto di Fisica Teorica, Naples, Italy

AND

G. BICZÓ
Central Research Institute for Chemistry of the Hungarian Academy of Sciences, Budapest, Hungary
(Received 5 August 1966)

A general self-consistent-field tight-binding linear-combination-of-atomic-orbitals (LCAO) formalism is given for three-dimensional polymers containing many atoms in the elementary cell with all neighbors interacting, taking overlap explicitly into account. This formalism, which corresponds essentially to the formulation given by Roothaan for closed-shell molecules, has been developed with the aid of Hermitian complex matrices. The special cases of nearest-neighbor approximation and of a linear chain are then derived from the general expression obtained. Finally, formulas are given, again in complex-matrix formulation, for the dependence of the energy levels and wave functions of the polymer on the wave number k.

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 40, NUMBER 9 1 MAY 1964

Semiempirical SCF–LCAO–MO Calculation of the Electronic Structure of the Guanine–Cytosine Base Pair: Possible Interpretation of the Mutagenic Effect of Radiation*

ROBERT REIN AND JANOS LADIK†

Quantum Chemistry Group for Research in Atomic, Molecular and Solid-State Theory, Uppsala University, Uppsala, Sweden

(Received 2 December 1963)

The π-electronic structure of the guanine–cytosine nucleotide base pair has been investigated in the semiempirical SCF–LCAO–MO approximation.

An examination of the energy levels with respect to the single bases showed that the transition between the highest filled and lowest unfilled π orbitals ($n^* \rightarrow n^*+1$) takes place between the levels originating from two different single bases. This statement follows from considering the n^* and n^*+1 levels of the base pair as perturbed ones of the guanine, cytosine single bases, respectively. This view is substantiated by the structure of the eigenvectors of the respective states. The transition between these states is associated with a transfer of charge, which in the investigated case involves the transfer of a unit charge from guanine to cytosine.

Since in the charge-transfer state the bases are essentially in an ionic form, a considerable increase in the probability of a single-proton tunnelling, proposed by Löwdin, may be expected. The interpretation of radiation-induced mutations in terms of the charge-transfer state and single-proton tunneling is further discussed.

Cooperativity in intermolecular forces

Theoret. Chim. Acta (Berl.) 34, 115—127 (1974)

Hydrogen Bonding in Long Chains of Hydrogen Fluoride and Long Chains and Large Clusters of Water Molecules

X. LCAO–MO Studies on Molecular Structure

Alfred Karpfen* and Janos Ladik
Lehrstuhl für Theoretische Chemie der Technischen Universität München

Peter Russegger and Peter Schuster**
Institut für Theoretische Chemie der Universität Wien

Sàndor Suhai
Central Research Institute for Chemistry of the Hungarian Academy of Sciences, Budapest

Received November 27, 1973

Energy band structures of one-dimensional $(HF)_n$- and $(H_2O)_n$-chains have been calculated (1) by extrapolation of CNDO/2–MO levels to infinite chain length and (2) by the CNDO/2 crystal orbital (CO) method. In the CO-calculations interactions up to fifth neighbours have been taken into account. Both types of calculations were performed using experimental geometries and CNDO/2 minimum geometries of the corresponding dimers $(HF)_2$ and $(H_2O)_2$. With the same geometries CO calculations on two-dimensional sheets of hydrogen bonded chains were performed too.

Due to end-effects the extrapolated MO bands are much broader than the bands obtained by the CO method. In the CO calculations further neighbour interactions play a non-negligible role and hence the nearest neighbour approximation is not sufficient for an accurate description of crystals containing hydrogen bonds.

MO calculations on one-dimensional chains of both systems show that the hydrogen bond energies increase with the number of monomers indicating the presence of cooperative effects. The hydrogen bond energies calculated with the CO method are usually somewhat larger than those extrapolated from the MO results. In three-dimensional networks of $(H_2O)_n$, however, the additional stabilization of clusters with respect to dimers is drastically diminished.

Key words: $(HF)_n$-chains – $(H_2O)_n$-chains – Hydrogen bond energies

Cited by 42 articles

Calculation of the hole mobilities of the three homopolynucleotides, poly(guanilic acid), poly(adenilic acid), and polythymidine in the presence of water and Na^+ ions

Attila Bende,[1,2] Ferenc Bogár,[2,3] Ferenc Beleznay,[2,4] and János Ladik[2,*]

[1]*Department of Molecular and Biomolecular Physics, National Institute for R and D of Isotopic and Molecular Technologies, Str. Donath 65-103, C.P. 700, Cluj Napoca RO-400293, Romania*

[2]*Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstrasse 3, D-91058, Erlangen, Germany*

[3]*Supramolecular and Nanostructured Marerials Research Group of the Hungarian Academy of Sciences, University of Szeged, Dóm tér 8., 6720, Szeged, Hungary*

[4]*Research Institute for Technical Physics and Material Science, Hungarian Academy of Sciences, H-1121 Budapest, Konkoly-Thege Miklós út 29-33, Hungary*

(Received 20 June 2008; published 29 December 2008)

Recent high resolution x-ray diffraction experiments have determined the structure of nucleosomes. In it 147 base pair long DNA B superhelix is wrapped around the eight nucleohistone proteins. They have found that there are many hydrogen-bonds (H-bonds) between the negative sites phosphate (PO_4^-) groups DNA, and first of all there is the positively charged lysine and arginine side chains of the histones. This means that there is a non-negligible charge transfer from DNA to the proteins causing a hole current in DNA and an electronic one in the proteins. If the relative positions of the two macromolecules change due to some external disturbances, the DNA moves away from the protein and can be read. If this happens simultaneously at several nucleosomes and at many places in chromatin (built up from the nucleosomes), undesired genetic information becomes readable. This final end can cause the occurrence of oncoproteins at an undesired time point which most probably disturbs the self-regulation of a differentiated cell. The connection of these chain of events with the initiation of cancer is obvious. To look into the details of these events we have used the detailed band structures of the four homopolynucleotides in the presence of water and natrium (Na^+) ions calculated previously with the help of the *ab initio* Hartree-Fock crystal orbital method. We have found that in the case of three homopolynucleotides the width of their valence band is broad enough (~10 times broader than the thermal energy at 300 K) for the application of the simple deformation potential approximation for transport calculations. With the help of this we have determined the hole mobilities at 300 K and 180 K of poly(guanilic acid), poly(adenilic acid), and polythimidine (polycytidine has a too narrow valence band for the application of the deformation potential method). The obtained mobilities are large enough to allow Bloch-type conduction in these systems. At the end of the paper we discuss briefly the possible mechanism of charge transport in aperiodic DNA as a combination of Bloch-type conduction, hopping, and tunneling.

DOI: 10.1103/PhysRevE.78.061923 PACS number(s): 87.15.ag, 82.39.Jn, 87.10.Vg

THE JOURNAL OF CHEMICAL PHYSICS **128**, 105101 (2008)

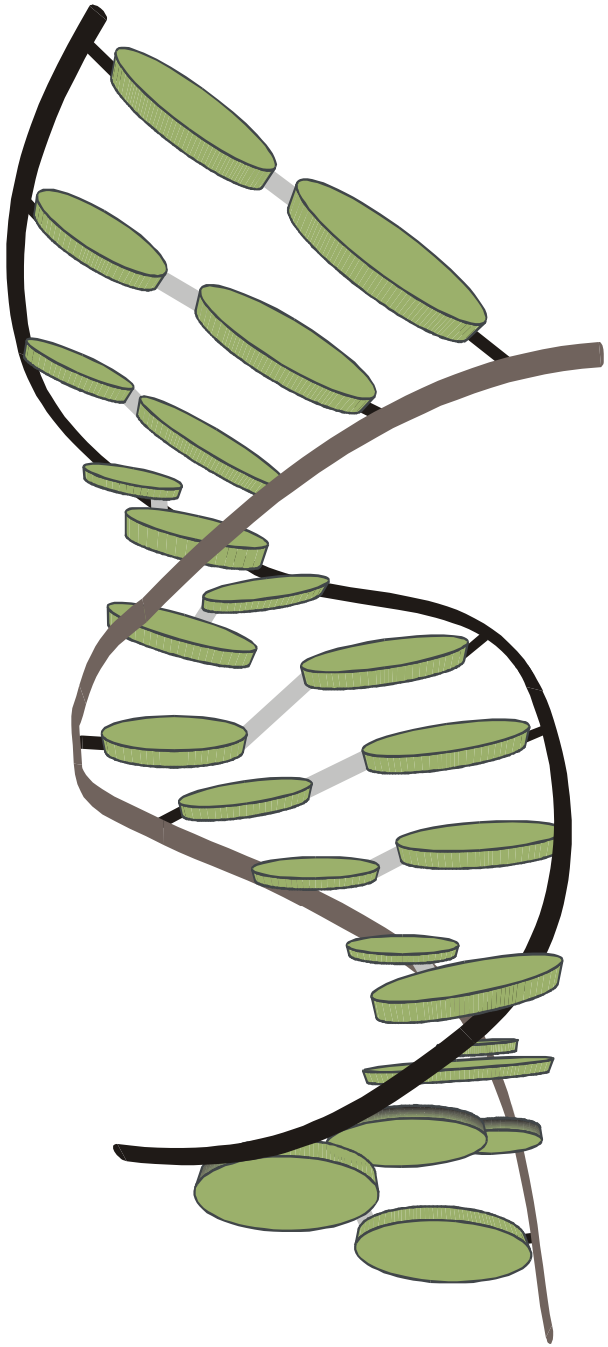
The electronic structure of the four nucleotide bases in DNA, of their stacks, and of their homopolynucleotides in the absence and presence of water

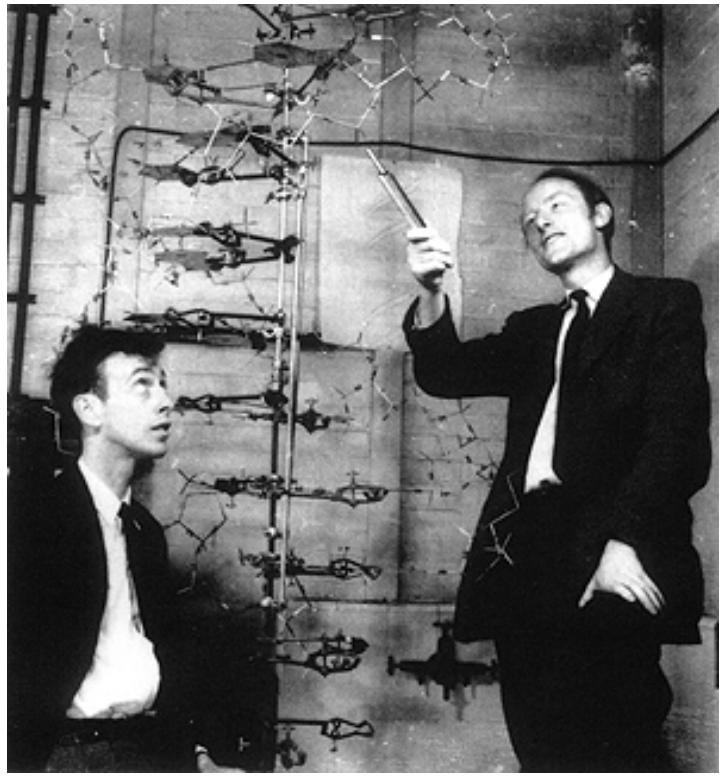
János Ladik,[1,a] Attila Bende,[1,2] and Ferenc Bogár[1,3]
[1]*Chair for Theoretical Chemistry and Laboratory of the National Foundation for Cancer Research, Friedrich-Alexander-University-Erlangen-Nürnberg, Egerlandstr. 3, 91058 Erlangen, Germany*
[2]*National Institute for Research and Development of Isotopic and Molecular Technologies, Str. Donath 71-103, C.P. 700, Cluj Napoca RO-400293, Romania*
[3]*Supramolecular and Nanostructured Materials Research Group of the Hungarian Academy of Sciences, University of Szeged, Dóm tér 8, 6720 Szeged, Hungary*

(Received 30 July 2007; accepted 14 December 2007; published online 10 March 2008)

Using the *ab initio* Hartree-Fock crystal orbital method in its linear combination of atomic orbital form, the energy band structure of the four homo-DNA-base stacks and those of poly(adenilic acid), polythymidine, and polycytidine were calculated both in the absence and presence of their surrounding water molecules. For these computations Clementi's double ζ basis set was applied. To facilitate the interpretation of the results, the calculations were supplemented by the calculations of the six narrow bands above the conduction band of poly(guanilic acid) with water. Further, the sugar-phosphate chain as well as the water structures around poly(adenilic acid) and polythymidine, respectively, were computed. Three important features have emerged from these calculations. (1) The nonbase-type or water-type bands in the fundamental gap are all close to the corresponding conduction bands. (2) The very broad conduction band (1.70 eV) of the guanine stack is split off to seven narrow bands in the case of poly(guanilic acid) (both without and with water) showing that in the energy range of the originally guanine-stack-type conduction band, states belonging to the sugar, to PO_4^-, to Na^+, and to water mix with the guanine-type states. (3) It is apparent that at the homopolynucleotides with water in three cases the valence bands are very similar (polycytidine, because it has a very narrow valence band, does not fall into this category). We have supplemented these calculations by the computation of correlation effects on the band structures of the base stacks by solving the inverse Dyson equation in its diagonal approximation taken for the self-energy the MP2 many body perturbation theory expression. In all cases the too large fundamental gap decreased by 2–3 eV. In most cases the widths of the valence and conduction bands, respectively, decreased (but not in all cases). This unusual behavior is most probably due to the rather large complexity of the systems. From all this emerges the following picture for the charge transport in DNA: There is a possibility in short segments of the DNA helix of a Bloch-type conduction of holes through the nucleotide base stacks of DNA combined with hopping (and in a lesser degree with tunneling). The motivation of this large scale computation was that recently in Zürich (ETH) they have performed high resolution x-ray diffraction experiments on the structure of the nucleosomes. The 8 nucleohistones in them are wrapped around by a DNA superhelix of 147 base pairs in the DNA B form. The most recent investigations have shown that between the DNA superhelix (mostly from its PO_4^- groups) there is a charge transfer to the positively charged side chains (first of all arginines and lysines) of the histones at 120 sites of the superhelix. This would cause a hole conduction in DNA and an electronic one in the proteins. © *2008 American Institute of Physics*.
[DOI: 10.1063/1.2832860]

DNA, the ‚star' among the biomolecules



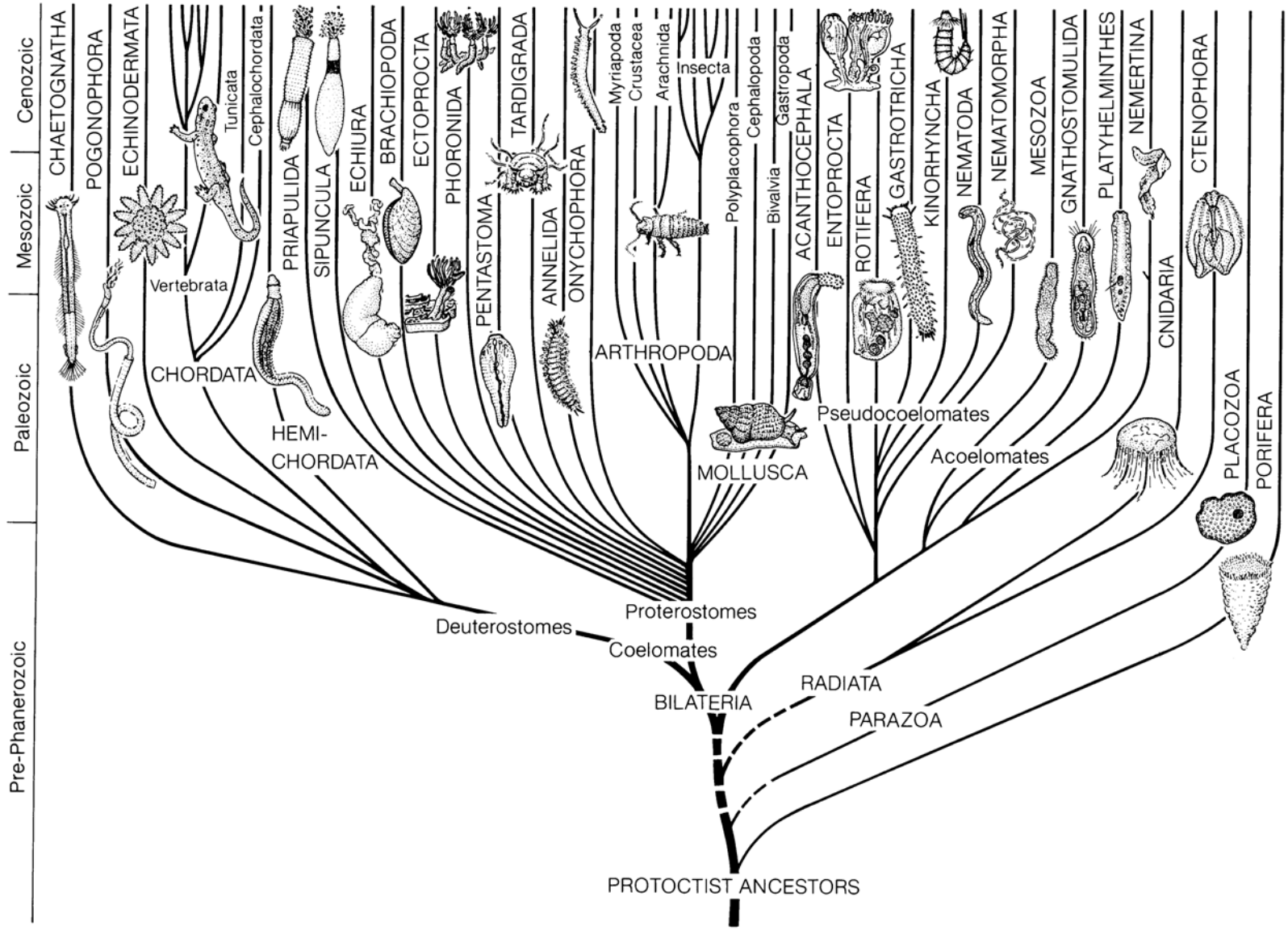


James D. Watson, 1928-, and Francis H.C. Crick, 1916-2004

Nobel prize 1962

1953 – 2003 fifty years double helix

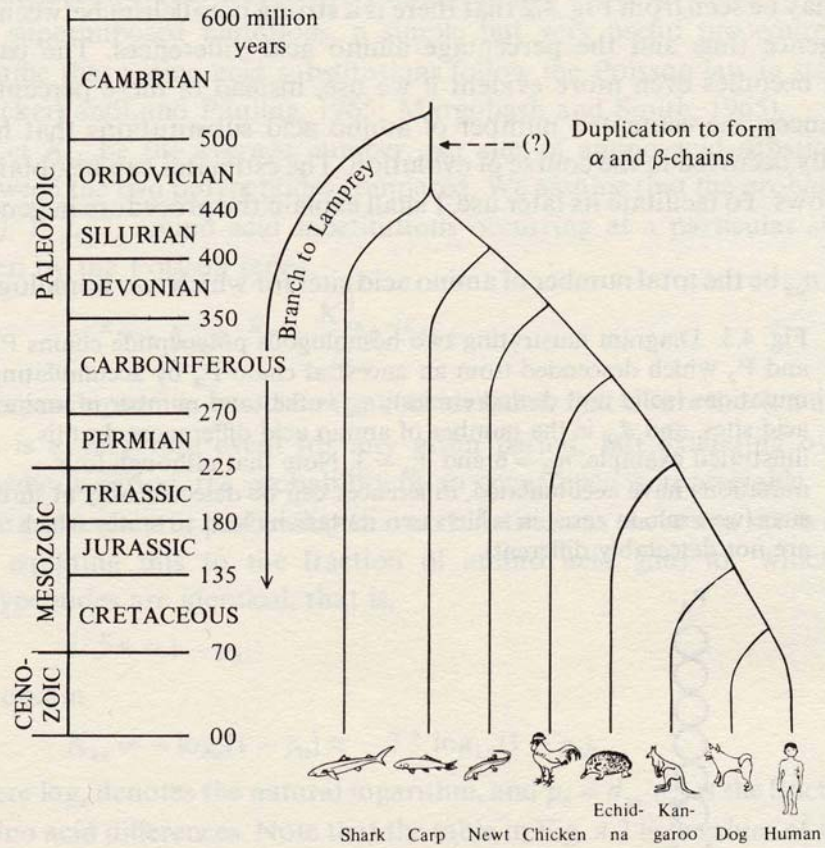
The three-dimensional structure of a short double helical stack of B-DNA



Modern phylogenetic tree: Lynn Margulis, Karlene V. Schwartz. *Five Kingdoms. An Illustrated Guide to the Phyla of Life on Earth.* W.H. Freeman, San Francisco, 1982.

The molecular clock of evolution

Fig. 4.2. Percentage amino acid differences when the α hemoglobin chains are compared among eight vertebrates together with their phylogenetic relationship and the times of divergence.

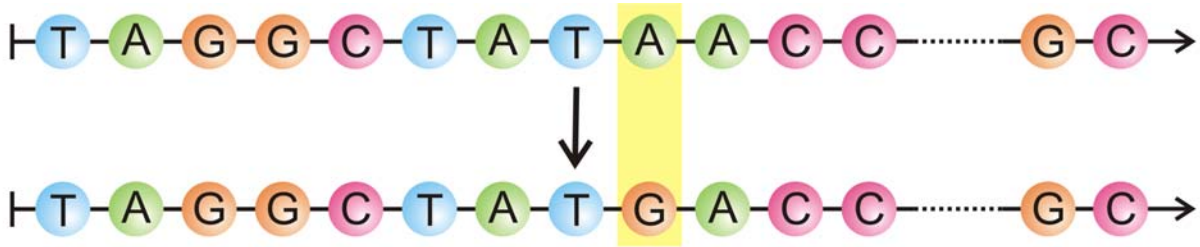


	Shark	Carp	Newt	Chicken	Echidna	Kangaroo	Dog	Human
Shark		59.4	61.4	59.7	60.4	55.4	56.8	53.2
Carp			53.2	51.4	53.6	50.7	47.9	48.6
Newt				44.7	50.4	47.5	46.1	44.0
Chicken					34.0	29.1	31.2	24.8
Echidna						34.8	29.8	26.2
Kangaroo							23.4	19.1
Dog								16.3
Human								

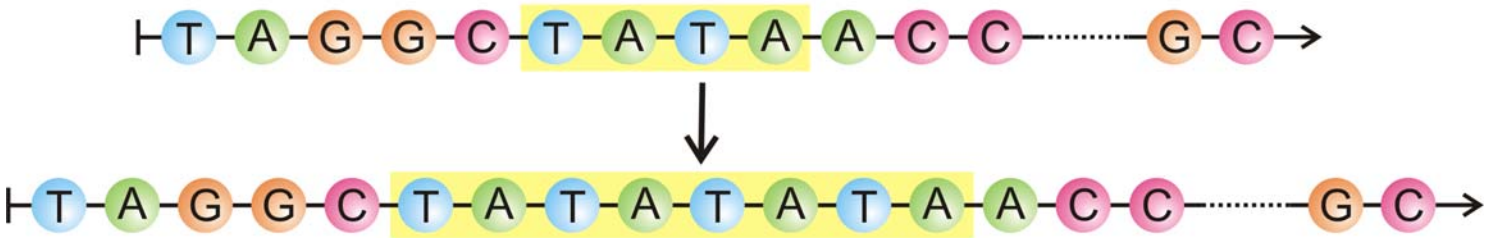
α-chain

Percent Amino Acid Differences

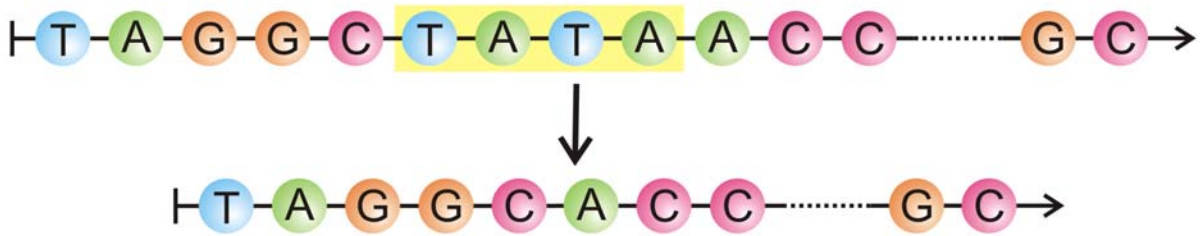
Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press. Cambridge, UK, 1983.



Point mutation

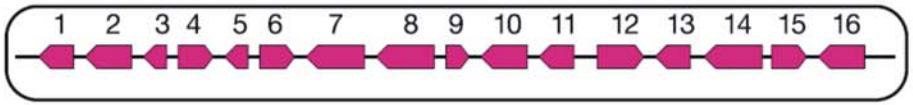


Insertion



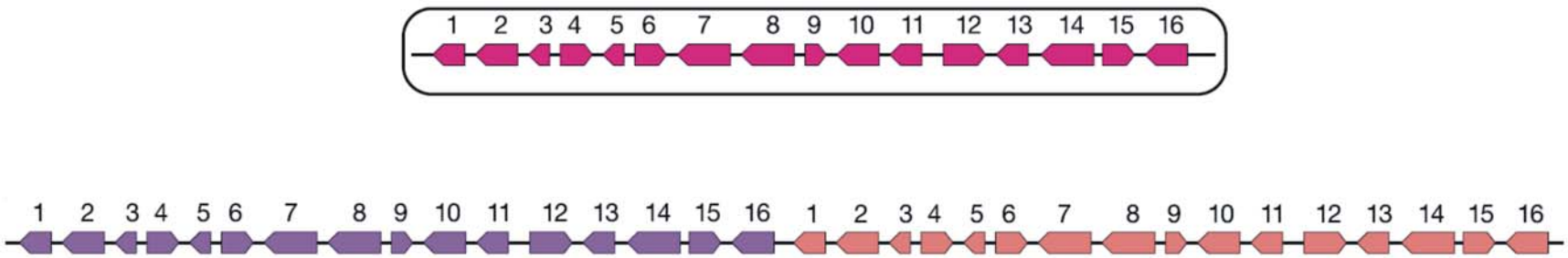
Deletion

Reconstruction of phylogenies through comparison of molecular sequence data



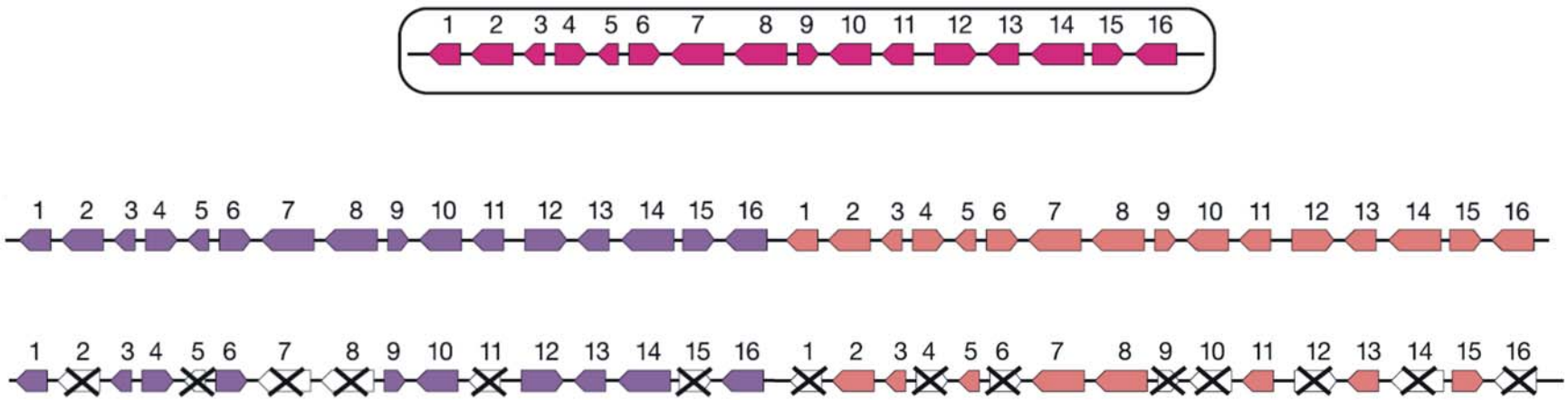
A model for the genome duplication in yeast 100 million years ago

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



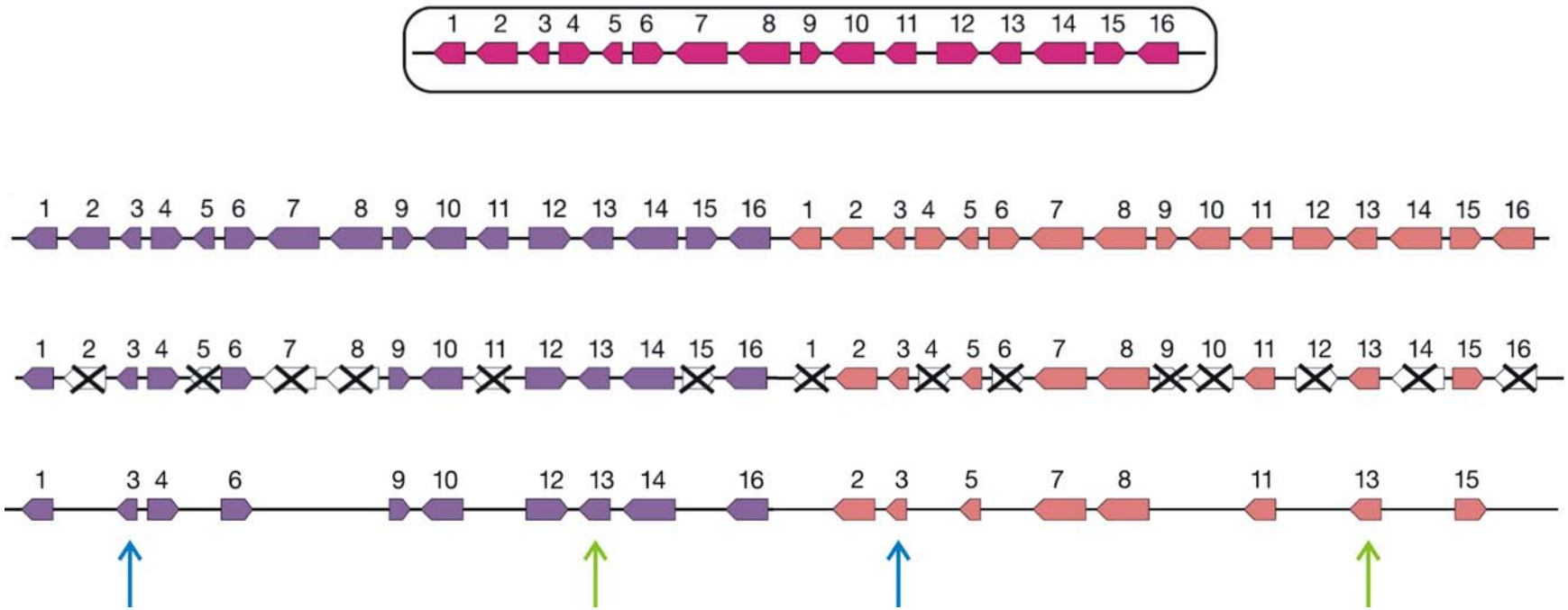
A model for the genome duplication in yeast 100 million years ago

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



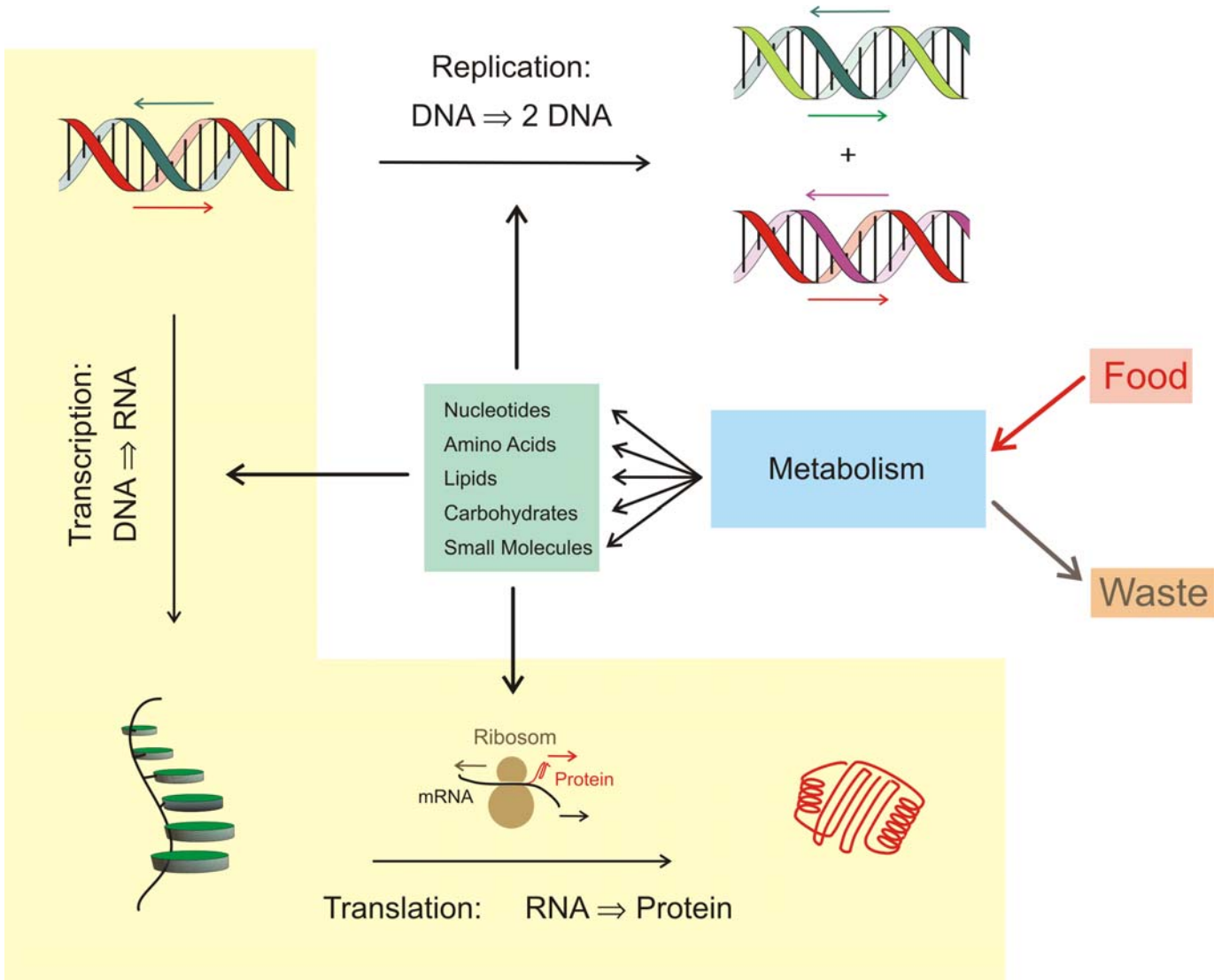
A model for the genome duplication in yeast 100 million years ago

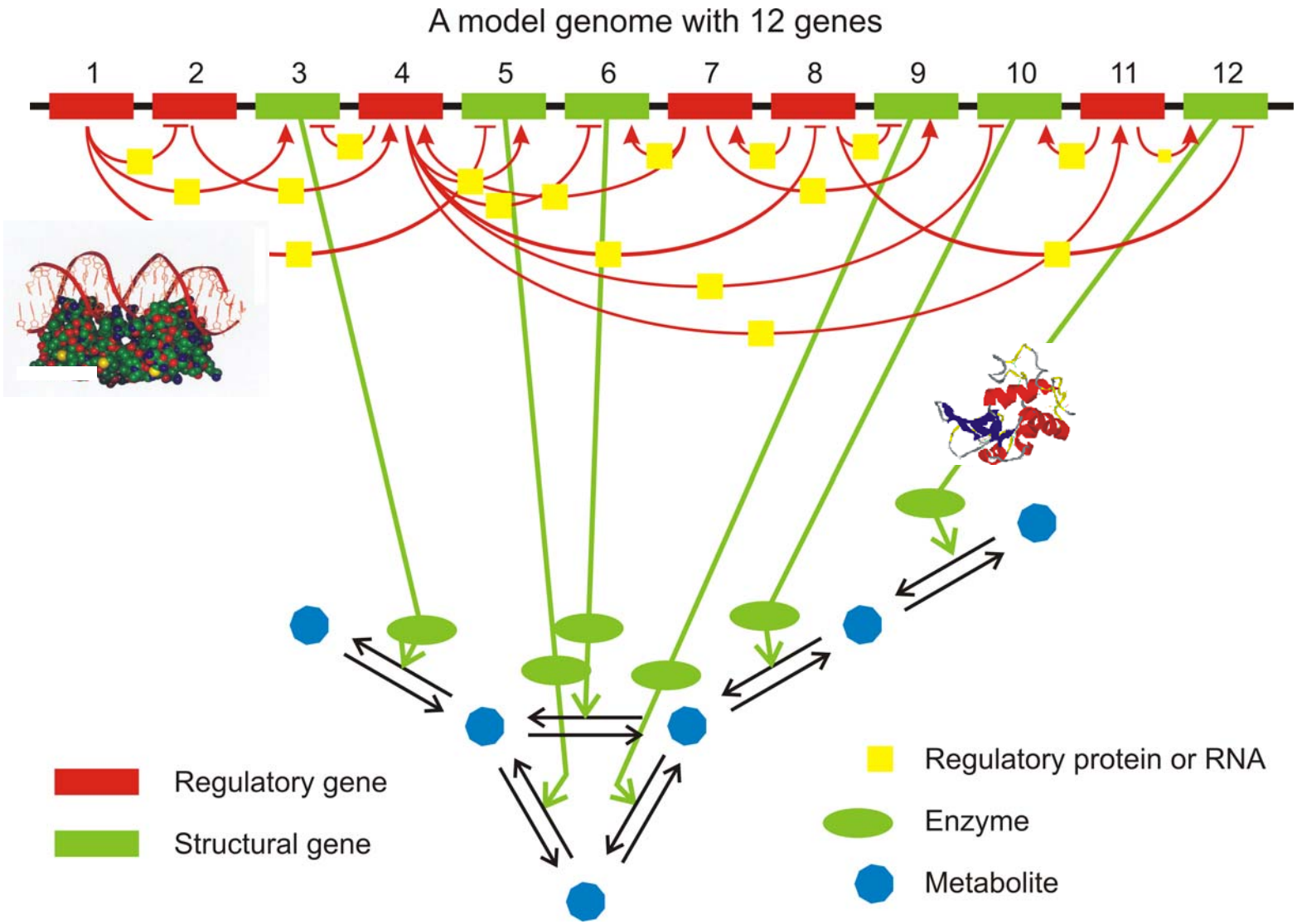
Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004



A model for the genome duplication in yeast 100 million years ago

Manolis Kellis, Bruce W. Birren, and Eric S. Lander. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624, 2004


Replication:
DNA ⇒ 2 DNA
+
Transcription:
DNA ⇒ RNA
Nucleotides
Amino Acids
Lipids
Carbohydrates
Small Molecules
Metabolism
Food
Waste
Ribosom
Protein
mRNA
Translation: RNA ⇒ Protein



A sketch of a genetic and metabolic network

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 246 (2007) 395–419
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Dynamic patterns of gene regulation I: Simple two-gene systems

Stefanie Widder[a], Josef Schicho[b], Peter Schuster[a,c,*]

[a]*Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria*
[b]*RICAM—Johann Radon Institute for Computational and Applied Mathematics of the Austrian Academy of Sciences, Altenbergerstraße 69, A-4040 Linz, Austria*
[c]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*

Received 24 February 2006; received in revised form 7 January 2007; accepted 8 January 2007
Available online 16 January 2007

Abstract

Regulation of gene activities is studied by means of computer assisted mathematical analysis of ordinary differential equations (ODEs) derived from binding equilibria and chemical reaction kinetics. Here, we present results on cross-regulation of two genes through activator and/or repressor binding. Arbitrary (differentiable) binding function can be used but systematic investigations are presented for gene–regulator complexes with integer valued Hill coefficients up to $n = 4$. The dynamics of gene regulation is derived from bifurcation patterns of the underlying systems of kinetic ODEs. In particular, we present analytical expressions for the parameter values at which one-dimensional (transcritical, saddle-node or pitchfork) and/or two-dimensional (Hopf) bifurcations occur. A classification of regulatory states is introduced, which makes use of the sign of a 'regulatory determinant' D (being the determinant of the block in the Jacobian matrix that contains the derivatives of the regulator binding functions): (i) systems with $D<0$, observed, for example, if both proteins are activators or repressors, to give rise to one-dimensional bifurcations only and lead to bistability for $n \geqslant 2$ and (ii) systems with $D>0$, found for combinations of activation and repression, sustain a Hopf bifurcation and undamped oscillations for $n>2$. The influence of basal transcription activity on the bifurcation patterns is described. Binding of multiple subunits can lead to richer dynamics than pure activation or repression states if intermediates between the unbound state and the fully saturated DNA initiate transcription. Then, the regulatory determinant D can adopt both signs, plus and minus.

Keywords: Basal transcription; Bifurcation analysis; Cooperative binding; Gene regulation; Hill coefficient; Hopf bifurcation

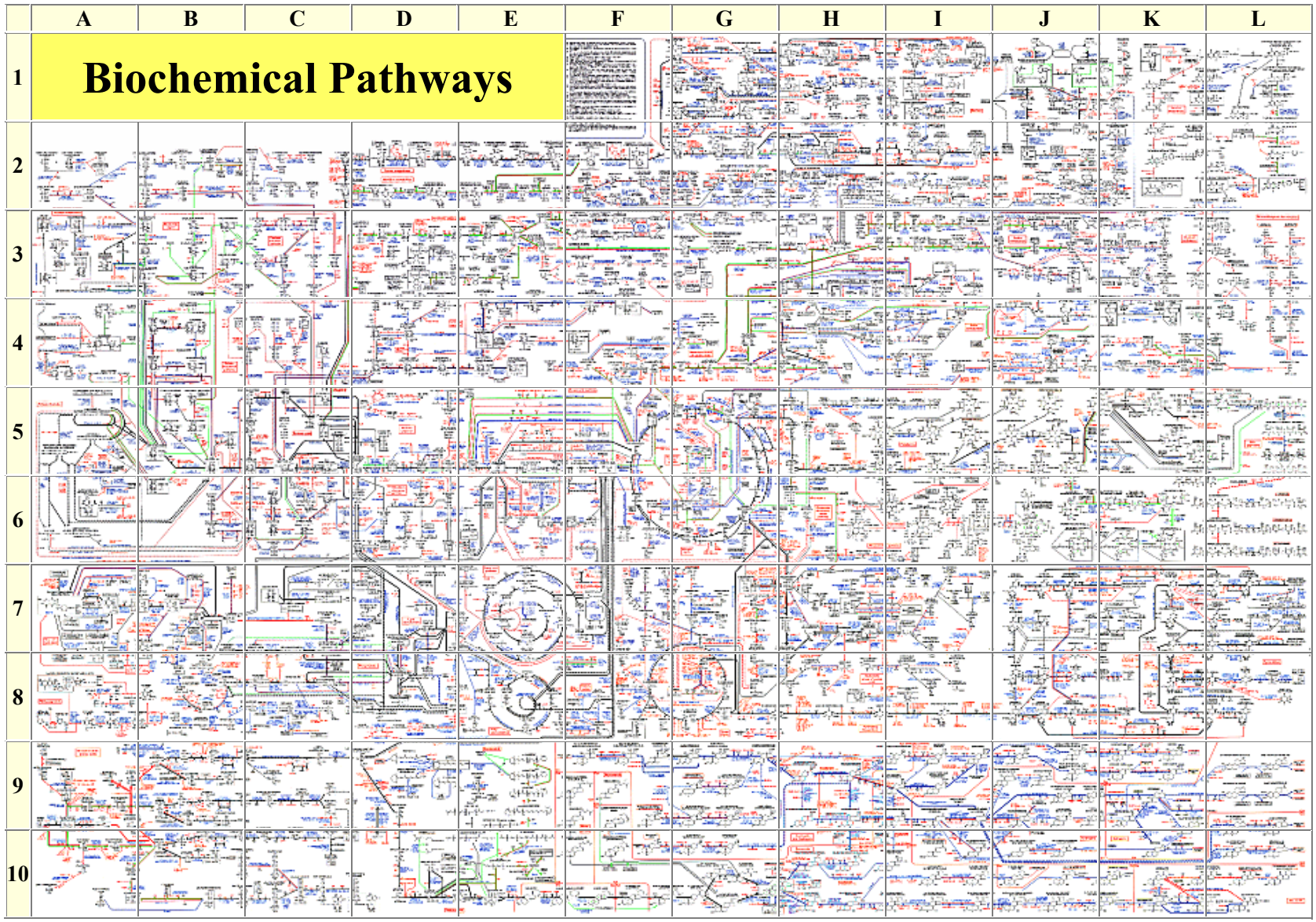
1. Introduction

Theoretical work on gene regulation goes back to the 1960s (Monod et al., 1963) soon after the first repressor protein had been discovered (Jacob and Monod, 1961). A little later the first paper on oscillatory states in gene regulation was published (Goodwin, 1965). The interest in gene regulation and its mathematical analysis never ceased (Tiwari et al., 1974; Tyson and Othmer, 1978; Smith, 1987) and saw a great variety of different attempts to design models of genetic regulatory networks that can be used in systems biology for computer simulation of *gen*(etic and met)*abolic* networks.[1] Most models in the literature aim at a minimalist dynamic description which, nevertheless, tries to account for the basic regulatory functions of large networks in the cell in order to provide a better understanding of cellular dynamics. A classic in general regulatory dynamics is the monograph by Thomas and D'Ari (1990). The currently used mathematical methods comprise application of Boolean logic (Thomas and Kaufman, 2001b; Savageau, 2001; Albert and Othmer, 2003), stochastic processes (Hume, 2000) and deterministic dynamic models, examples are Cherry and Adler (2000), Bindschadler and Sneyd (2001) and Kobayashi et al. (2003) and the recent elegant analysis of bistability (Craciun et al.,

*Corresponding author. Institut für Theoretische Chemie der Universität Wien, Währingerstraße 17, A-1090 Wien, Austria.
Tel.: +43 1 4277 52743; fax: +43 1 4277 52793.
E-mail address: pks@tbi.univie.ac.at (P. Schuster).

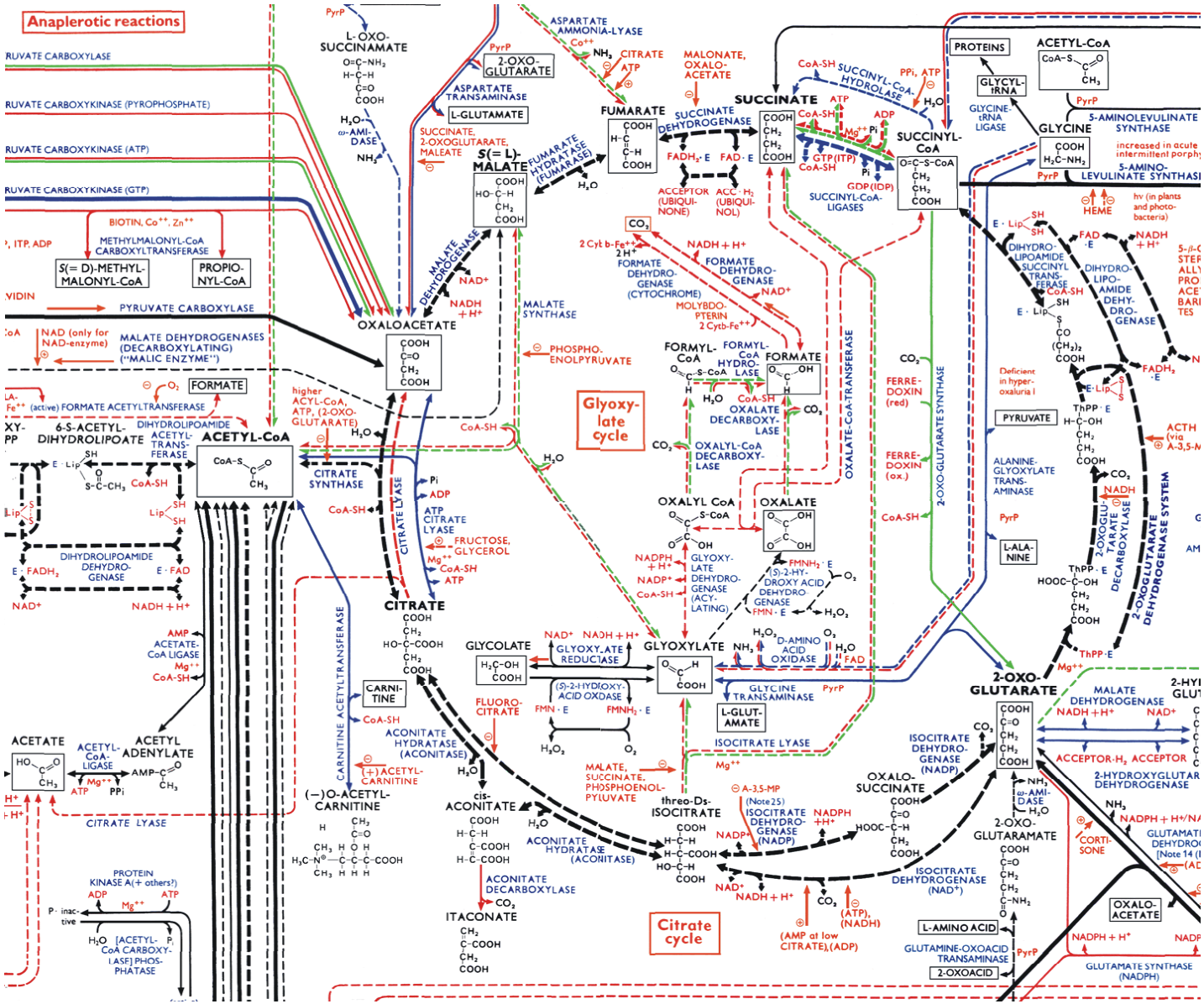
[1]Discussion and analysis of combined genetic and metabolic networks has become so frequent and intense that we suggest to use a separate term, *genabolic networks*, for this class of complex dynamical systems.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.01.004



The reaction network of cellular metabolism published by Boehringer-Ingelheim.

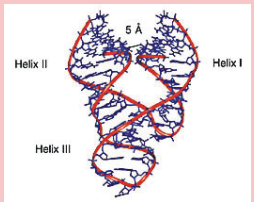
The citric acid or Krebs cycle (enlarged from previous slide).



RNA, the ‚magic' biomolecule

RNA

RNA *as catalyst*



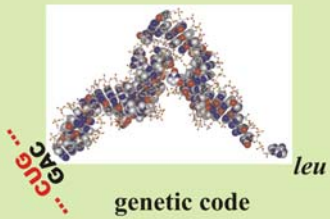
Ribozyme

RNA *as transmitter of genetic information*

DNA
↓ transcription
...AGAGCGCCAGACUGAAGAUCUGGAGGUCCUGUGUUC...
messenger-RNA
↓ translation
protein

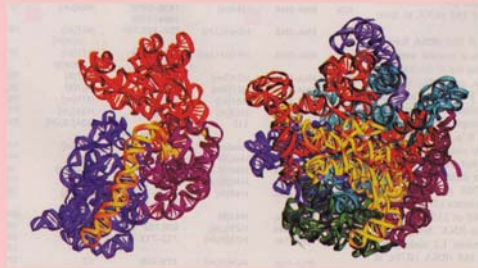
RNA as **working copy** of genetic information

RNA *as adapter molecule*



genetic code

RNA *is the catalytic subunit in supramolecular complexes*

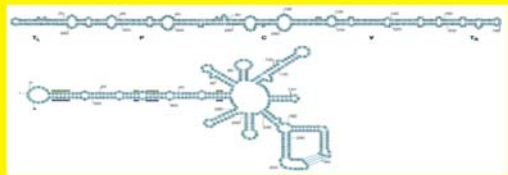


The **ribosome** is a **ribozyme !**

The* RNA *world as a precursor of the current* DNA + protein *biology

RNA *is modified by epigenetic control*

RNA editing, alternative splicing



Viroids

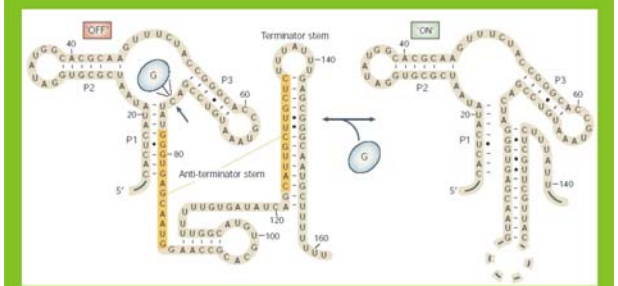
RNA *as carrier of genetic information*

RNA viruses and retroviruses

RNA evolution *in vitro*

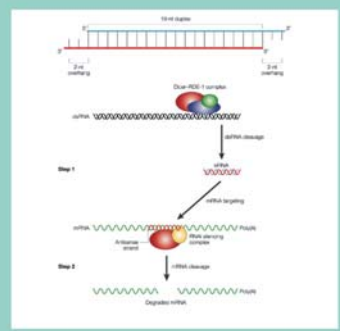
Evolutionary biotechnology
RNA aptamers, artificial ribozymes, allosteric ribozymes

***Allosteric control of transcribed* RNA**



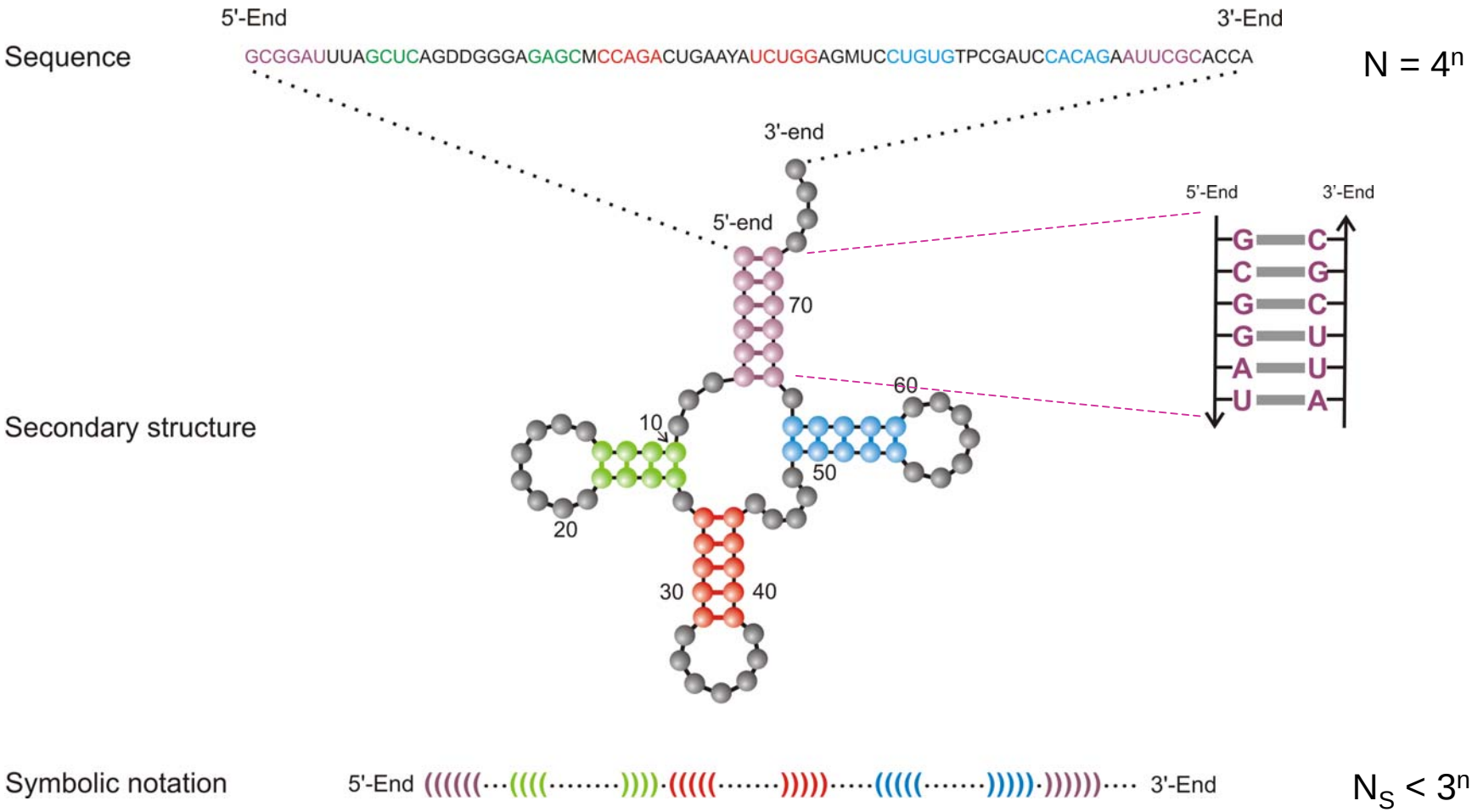
Riboswitches controlling transcription and translation through **metabolites**

RNA *as regulator of gene expression*



Gene regulation by small interfering RNAs

RNA – The magic molecule



Criterion: Minimum free energy (mfe)

Rules: _ (_) _ ∈ {**AU**,**CG**,**GC**,**GU**,**UA**,**UG**}

A symbolic notation of RNA secondary structure that is equivalent to the conventional graphs

The notion of RNA (secondary) structure

1. Minimum free energy structure
2. Many sequences one structure
3. Suboptimal structures
4. Kinetic structures

The notion of RNA (secondary) structure

1. **Minimum free energy structure**
2. Many sequences one structure
3. Suboptimal structures
4. Kinetic structures

RNA sequence GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA folding:

Structural biology, spectroscopy of biomolecules, understanding **molecular function**

Biophysical chemistry: thermodynamics and kinetics

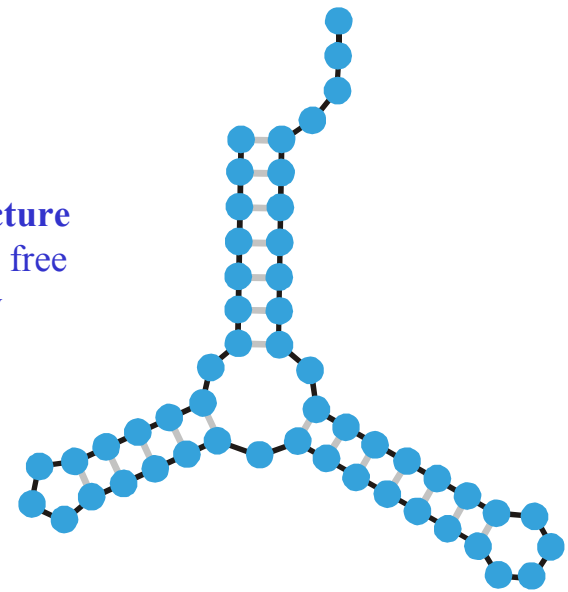
Empirical parameters

Vienna RNA-Package

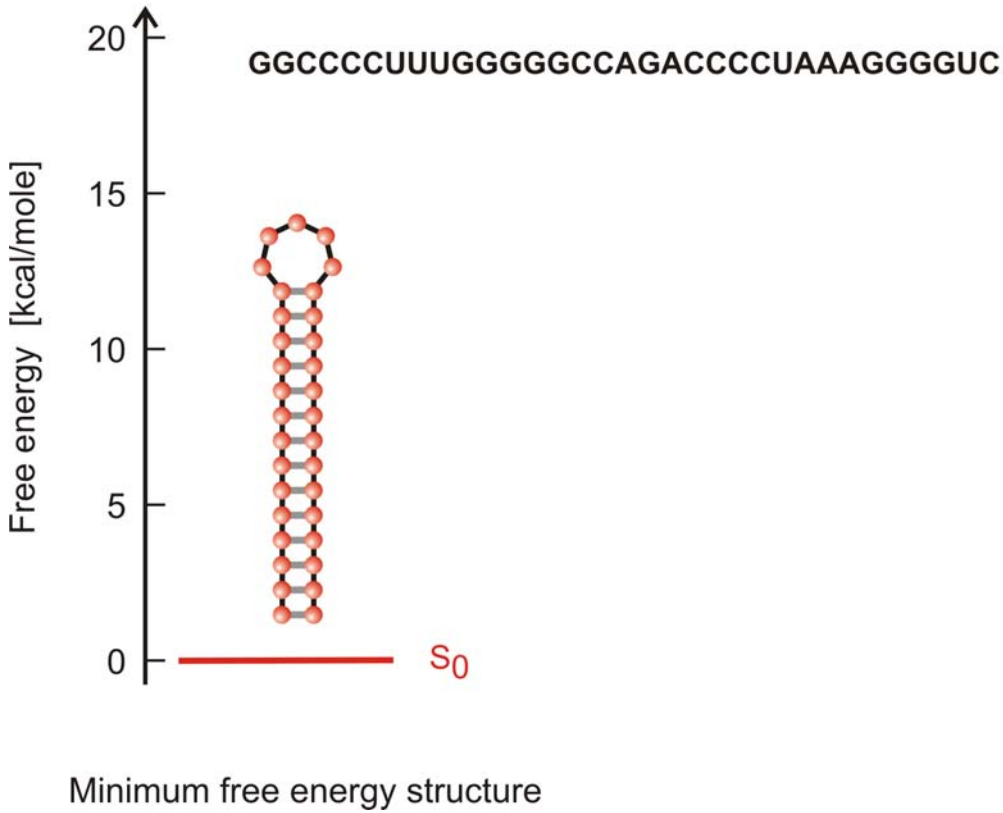
Version 1.8.3

`http://www.tbi.univie.ac.at`

RNA structure of minimal free energy



Sequence, structure, and design



Extension of the notion of structure

Monatshefte für Chemie 125, 167–188 (1994)

Monatshefte für Chemie
Chemical Monthly

Fast Folding and Comparison of RNA Secondary Structures

I. L. Hofacker[1,*], W. Fontana[3], P. F. Stadler[1,3], L. S. Bonhoeffer[4], M. Tacker[1] and P. Schuster[1,2,3]

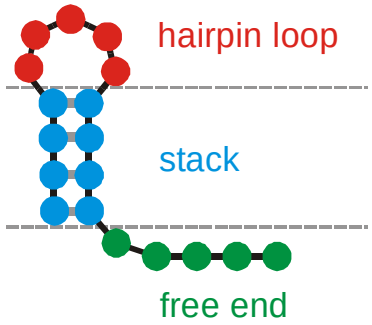
[1] Institut für Theoretische Chemie, Universität Wien, A-1090 Wien, Austria
[2] Institut für Molekulare Biotechnologie, D-07745 Jena, Federal Republic of Germany
[3] Santa Fe Institute, Santa Fe, NM 87501, U.S.A.
[4] Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

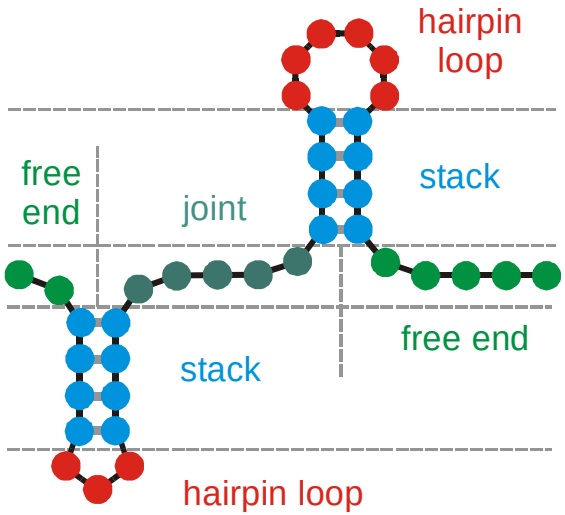
Summary. Computer codes for computation and comparison of RNA secondary structures, the Vienna RNA package, are presented, that are based on dynamic programming algorithms and aim at predictions of structures with minimum free energies as well as at computations of the equilibrium partition functions and base pairing probabilities.

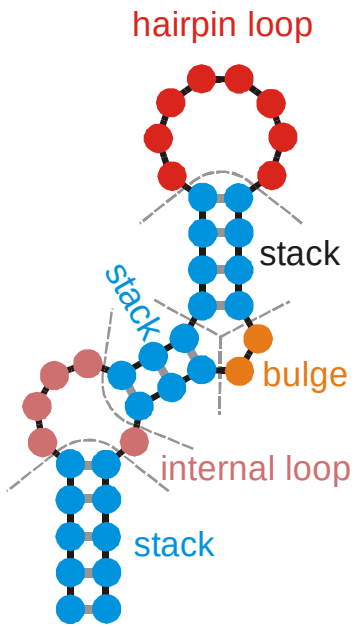
An efficient heuristic for the inverse folding problem of RNA is introduced. In addition we present compact and efficient programs for the comparison of RNA secondary structures based on tree editing and alignment.

All computer codes are written in ANSI C. They include implementations of modified algorithms on parallel computers with distributed memory. Performance analysis carried out on an Intel Hypercube shows that parallel computing becomes gradually more and more efficient the longer the sequences are.

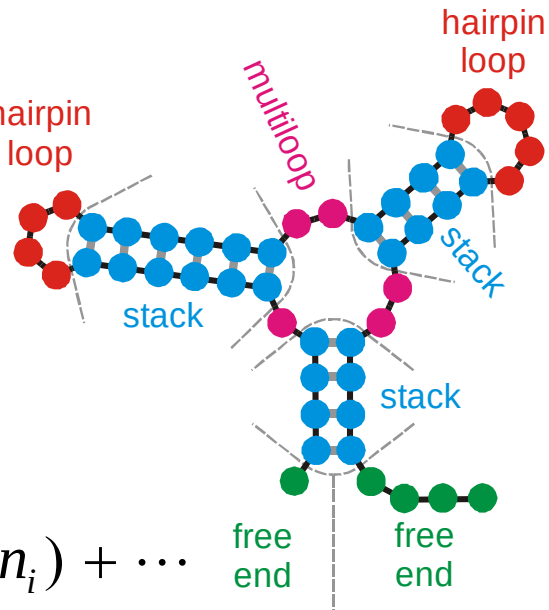
Keywords. Inverse folding; parallel computing; public domain software; RNA folding; RNA secondary structures; tree editing.





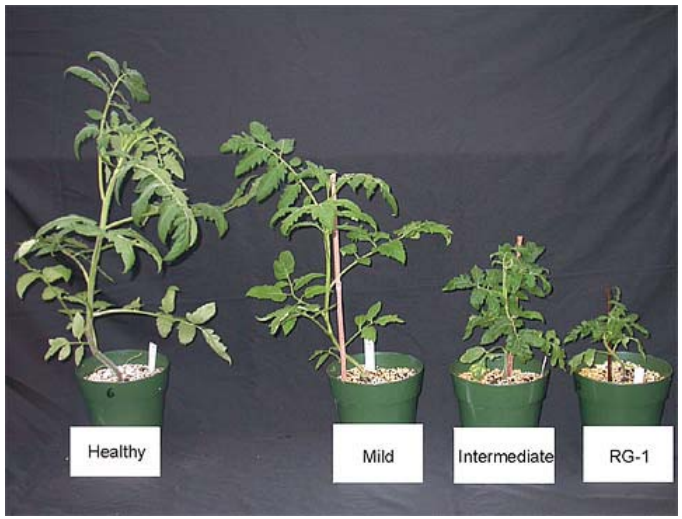


Elements of RNA secondary structures as used in free energy calculations



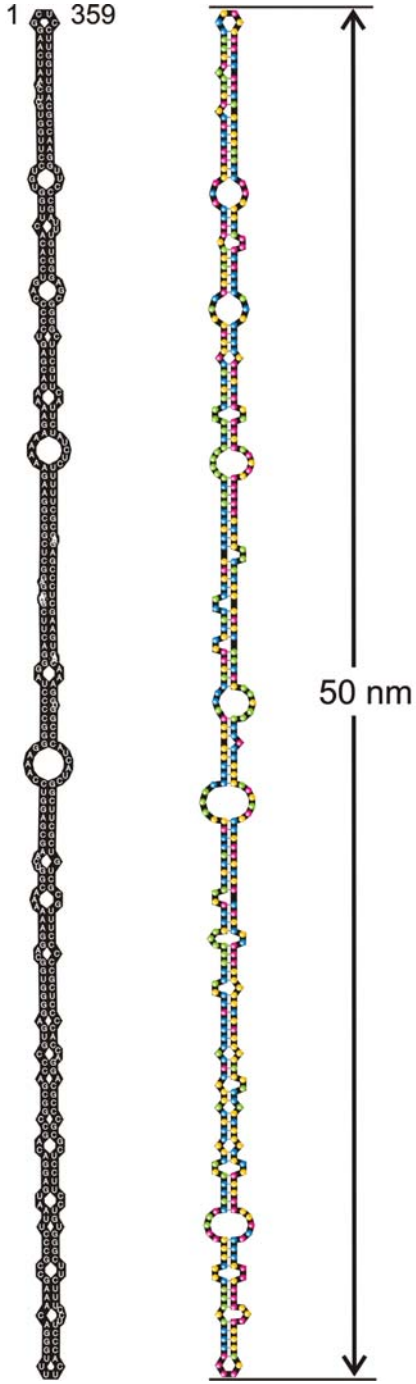
$$\Delta G_0^{300} = \underbrace{\sum g_{ij,kl}}_{\substack{\text{stacks of} \\ \text{base pairs}}} + \underbrace{\sum h(n_l)}_{\substack{\text{hairpin} \\ \text{loops}}} + \underbrace{\sum b(n_b)}_{\text{bulges}} + \underbrace{\sum i(n_i)}_{\substack{\text{internal} \\ \text{loops}}} + \cdots$$

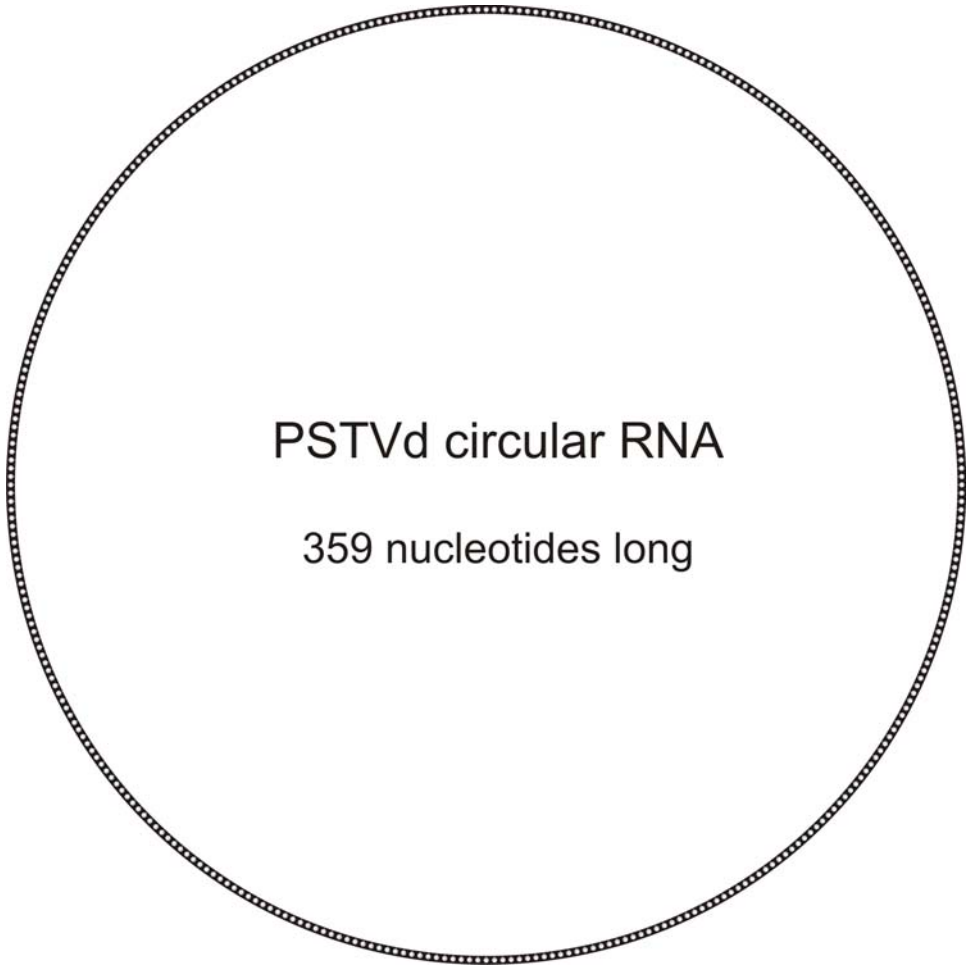
J. Demez. European and mediterranean plant protection organization archive. France



R.W. Hammond, R.A. Owens. Molecular Plant Pathology Laboratory, US Department of Agriculture

Plant damage by viroids

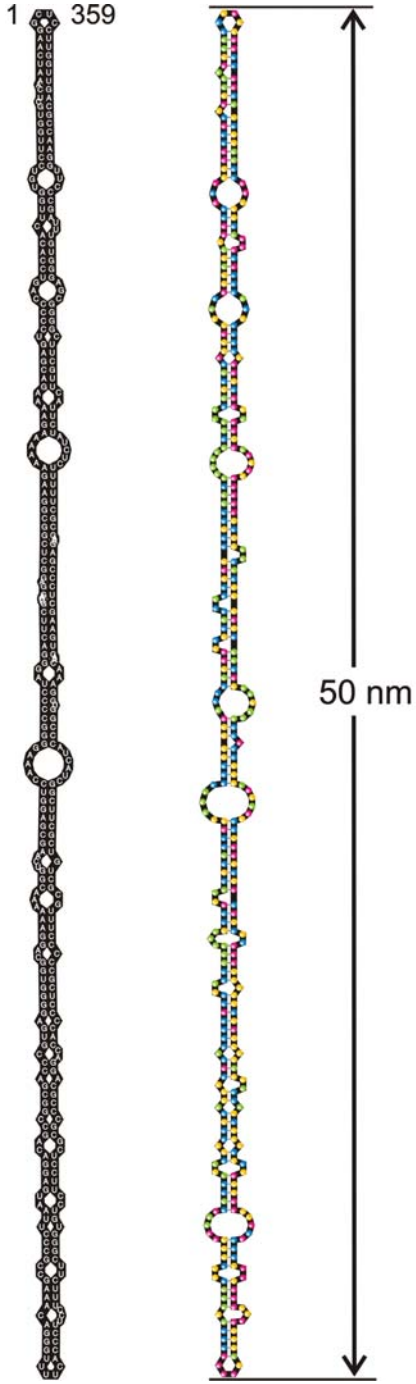




Adenine
Uracil
Guanine
Cytosine

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)



Vienna RNA Package 1.8.2

Biochemically supported structure

Nucleotide sequence and secondary structure of the potato spindle tuber viroid RNA

H.J.Gross, H. Domdey, C. Lossow, P Jank, M. Raba, H. Alberty, and H.L. Sänger. *Nature* **273**:203-208 (1978)

The notion of RNA (secondary) structure

1. Minimum free energy structure
2. **Many sequences one structure**
3. Suboptimal structures
4. Kinetic structures

RNA sequence GUAUCGAAAUACGUAGCGUAUGGGGAUGCUGGACGGUCCCAUCGGUACUCCA

RNA folding:

Structural biology, spectroscopy of biomolecules, understanding **molecular function**

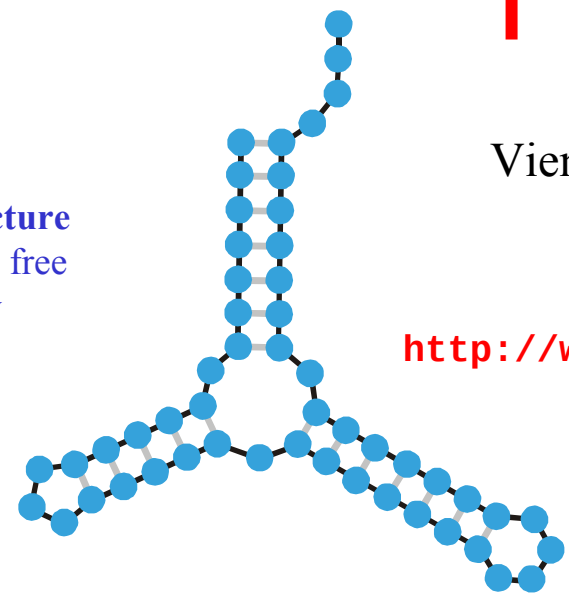
Iterative determination of a sequence for the given secondary structure

Inverse Folding Algorithm

Inverse folding of RNA:

Biotechnology, **design of biomolecules** with predefined structures and functions

RNA structure of minimal free energy



Vienna RNA-Package

Version 1.8.3

`http://www.tbi.univie.ac.at`

Sequence, structure, and design

Inverse folding algorithm

$\mathbf{I_0} \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow I_4 \rightarrow \ldots \rightarrow I_k \rightarrow I_{k+1} \rightarrow \ldots \rightarrow \mathbf{I_t}$

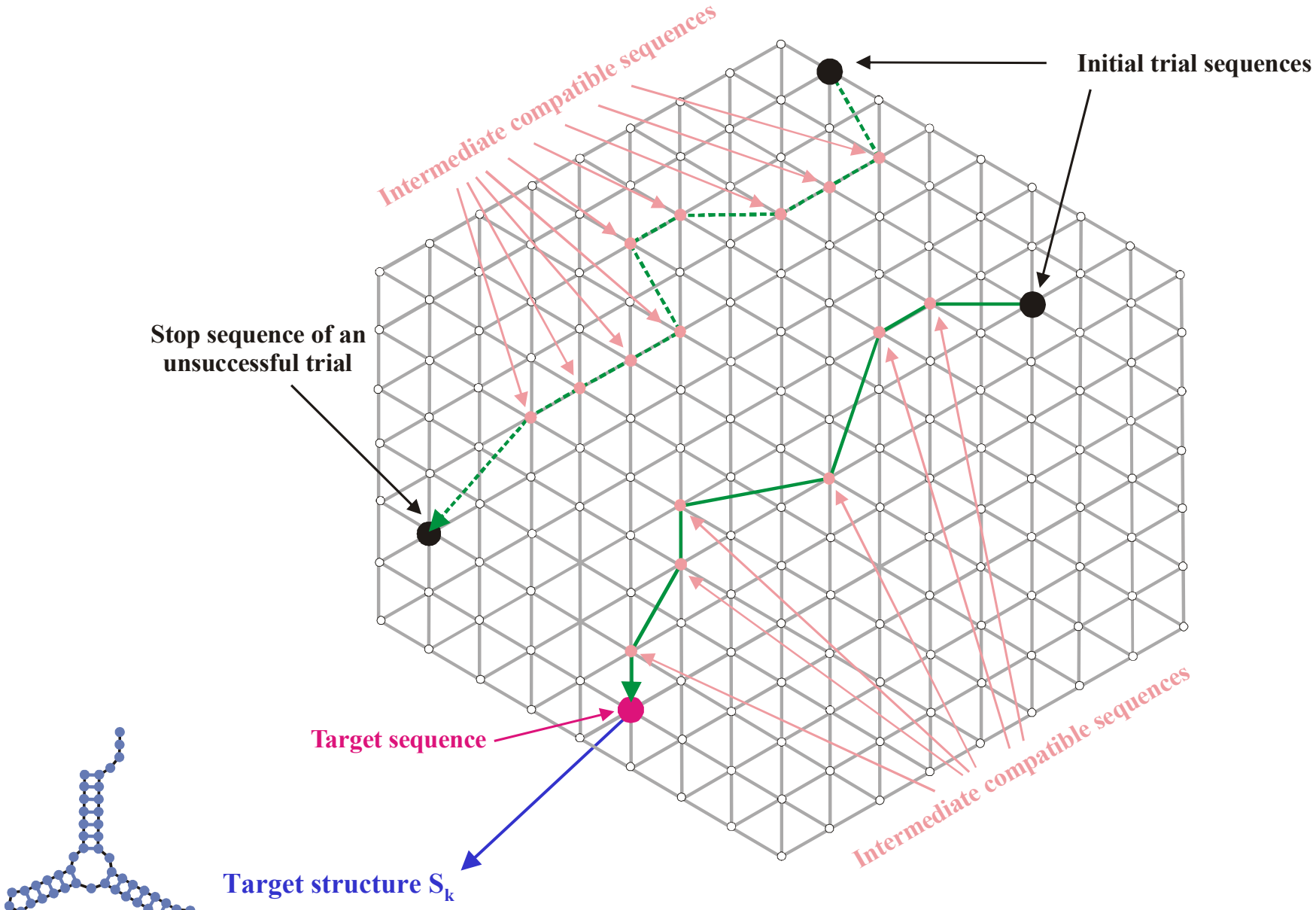
$\mathbf{S_0} \rightarrow S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4 \rightarrow \ldots \rightarrow S_k \rightarrow S_{k+1} \rightarrow \ldots \rightarrow \mathbf{S_t}$

$I_{k+1} = \mathfrak{M}_k(I_k)$ and $\Delta d_S(S_k,S_{k+1}) = d_S(S_{k+1},S_t) - d_S(S_k,S_t) < 0$

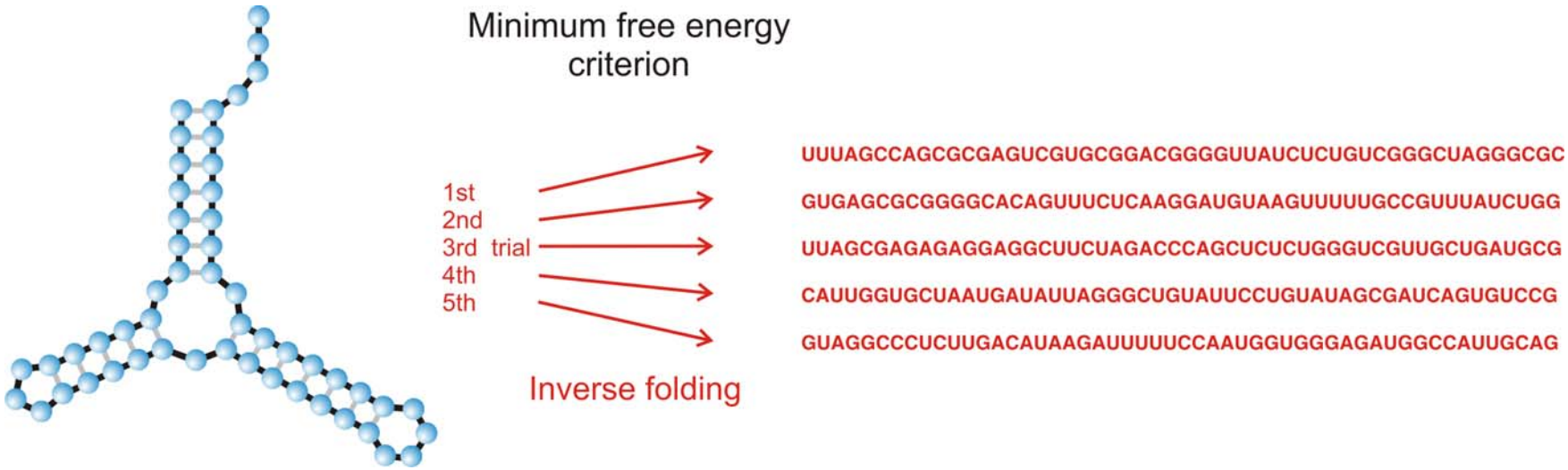
$\mathfrak{M}$ ... base or base pair mutation operator

$d_S\,(S_i,S_j)$ ... distance between the two structures S_i and S_j

‚Unsuccessful trial' ... termination after n steps



Approach to the **target structure S_k** in the inverse folding algorithm



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

From sequences to shapes and back: a case study in RNA secondary structures

PETER SCHUSTER[1,2,3], WALTER FONTANA[3], PETER F. STADLER[2,3] AND IVO L. HOFACKER[2]

[1] *Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, PF 100813, D-07708 Jena, Germany*
[2] *Institut für Theoretische Chemie, Universität Wien, Austria*
[3] *Santa Fe Institute, Santa Fe, U.S.A.*

SUMMARY

RNA folding is viewed here as a map assigning secondary structures to sequences. At fixed chain length the number of sequences far exceeds the number of structures. Frequencies of structures are highly non-uniform and follow a generalized form of Zipf's law: we find relatively few common and many rare ones. By using an algorithm for inverse folding, we show that sequences sharing the same structure are distributed randomly over sequence space. All common structures can be accessed from an arbitrary sequence by a number of mutations much smaller than the chain length. The sequence space is percolated by extensive neutral networks connecting nearest neighbours folding into identical structures. Implications for evolutionary adaptation and for applied molecular evolution are evident: finding a particular structure by mutation and selection is much simpler than expected and, even if catalytic activity should turn out to be sparse in the space of RNA structures, it can hardly be missed by evolutionary processes.

Proc. R. Soc. Lond. B (1994) **255**, 279–284
Printed in Great Britain

Space of genotypes: $\mathcal{I} = \{I_1, I_2, I_3, I_4, \ldots, I_N\}$; Hamming metric

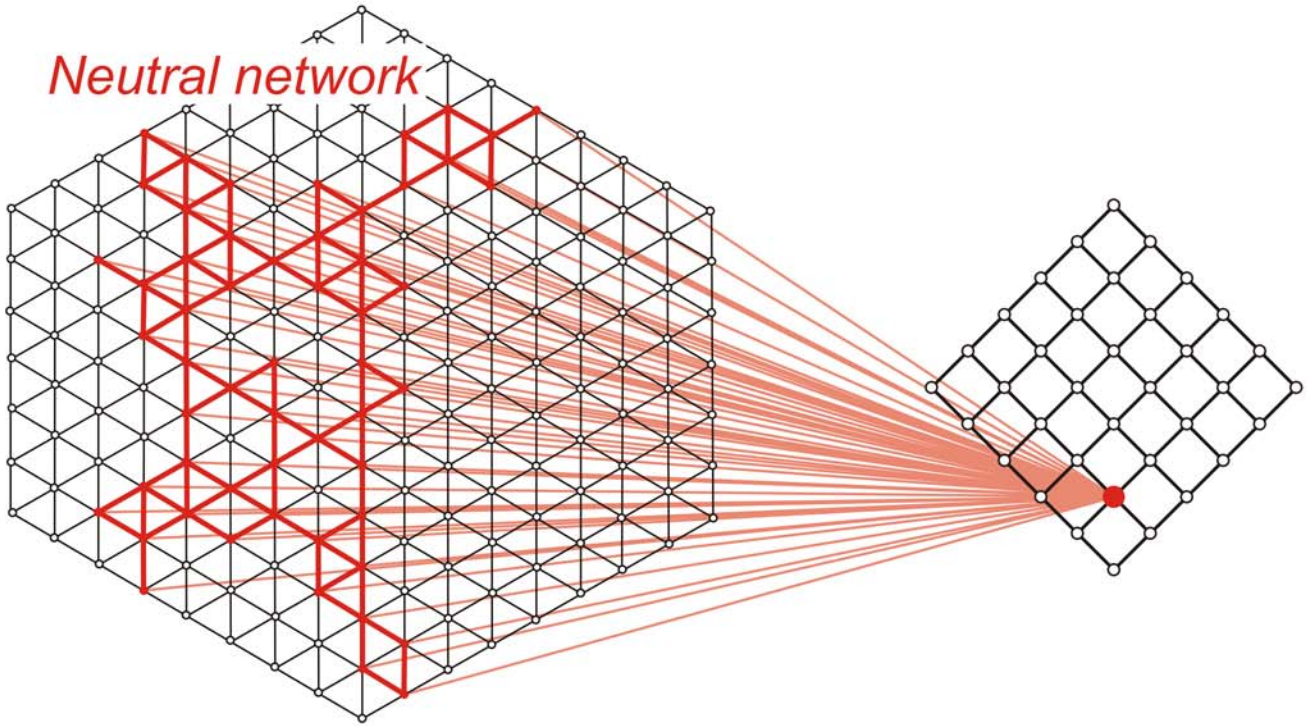
Space of phenotypes: $S = \{S_1, S_2, S_3, S_4, \ldots, S_M\}$; metric (not required)

$$N >> M$$

$$\psi(I_j) = S_k$$

$$\mathbf{G}_k = \psi^{-1}(S_k) \cup \{ I_j \mid \psi(I_j) = S_k \}$$

A mapping ψ and its inversion



Sequence space

Structure space

many genotypes ⇒ one phenotype

Bulletin of Mathematical Biology, Vol. 59, No. 2, pp. 339–397, 1997
Elsevier Science Inc.

0092-8240/97 $17.00 + 0.00

S0092-8240(96)00089-4

GENERIC PROPERTIES OF COMBINATORY MAPS: NEUTRAL NETWORKS OF RNA SECONDARY STRUCTURES[1]

■ CHRISTIAN REIDYS*,†, PETER F. STADLER*,‡ and PETER SCHUSTER*,‡,§,[2]
*Santa Fe Institute,
Santa Fe, NM 87501, U.S.A.

†Los Alamos National Laboratory,
Los Alamos, NM 87545, U.S.A.

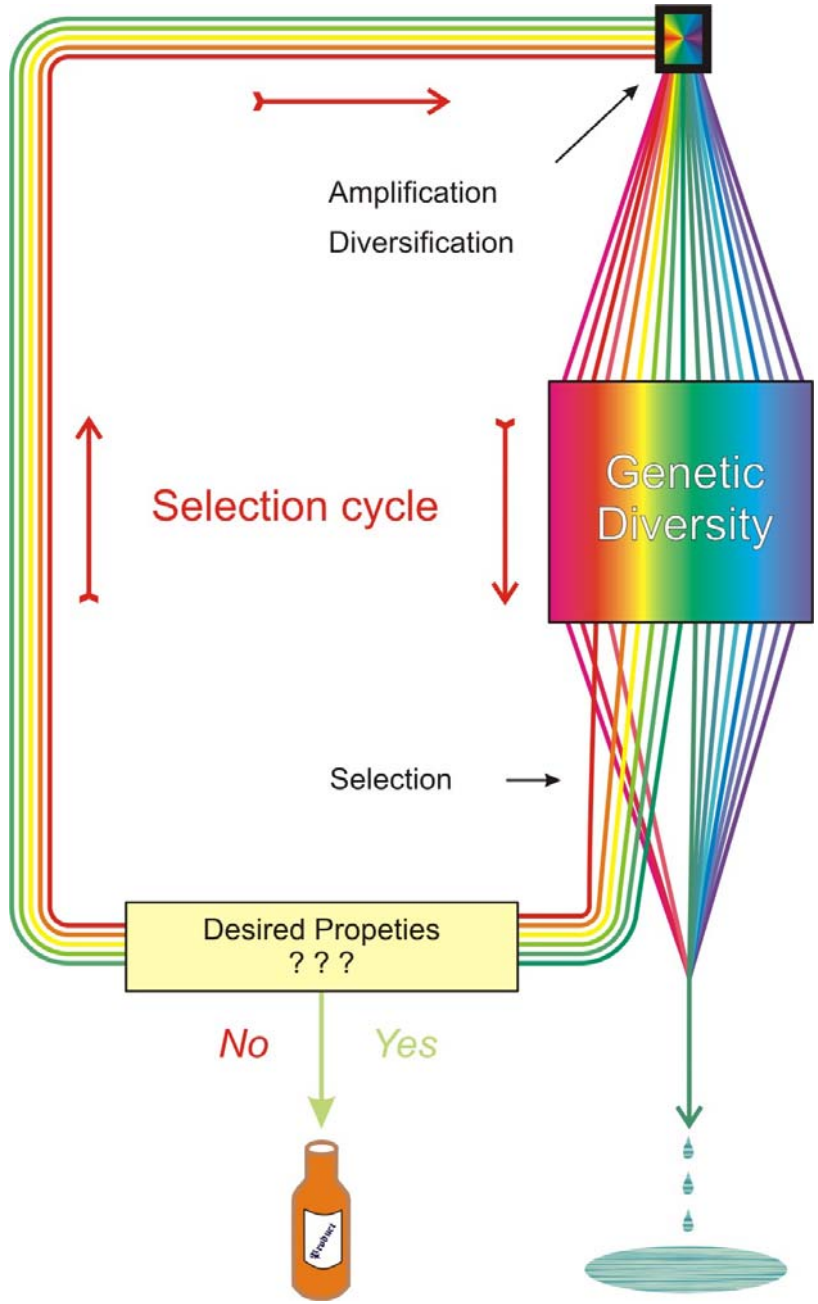
‡Institut für Theoretische Chemie der Universität Wien,
A-1090 Wien, Austria

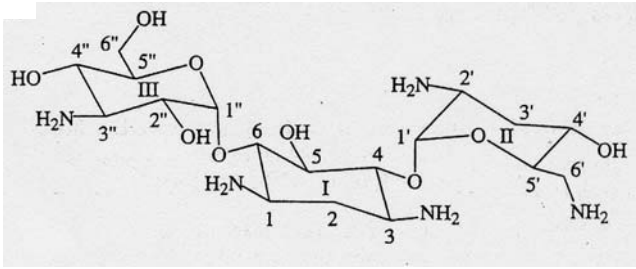
§Institut für Molekulare Biotechnologie,
D-07708 Jena, Germany

(*E.mail: pks@tbi.univie.ac.at*)

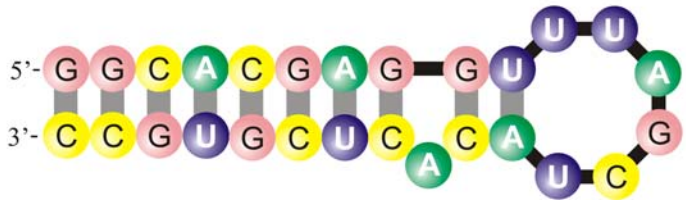
Random graph theory is used to model and analyse the relationships between sequences and secondary structures of RNA molecules, which are understood as mappings from sequence space into shape space. These maps are non-invertible since there are always many orders of magnitude more sequences than structures. Sequences folding into identical structures form *neutral networks*. A neutral network is embedded in the set of sequences that are *compatible* with the given structure. Networks are modeled as graphs and constructed by random choice of vertices from the space of compatible sequences. The theory characterizes neutral networks by the mean fraction of neutral neighbors (λ). The networks are connected and percolate sequence space if the fraction of neutral nearest neighbors exceeds a threshold value ($\lambda > \lambda^*$). Below threshold ($\lambda < \lambda^*$), the networks are partitioned into a largest "giant" component and several smaller components. Structures are classified as "common" or "rare" according to the sizes of their pre-images, i.e. according to the fractions of sequences folding into them. The neutral networks of any pair of two different common structures almost touch each other, and, as expressed by the conjecture of *shape space covering* sequences folding into almost all common structures, can be found in a small ball of an arbitrary location in sequence space. The results from random graph theory are compared to data obtained by folding large samples of RNA sequences. Differences are explained in terms of specific features of RNA molecular structures. © *1997 Society for Mathematical Biology*

An example of ‘artificial selection’ with RNA molecules or ‘breeding’ of biomolecules





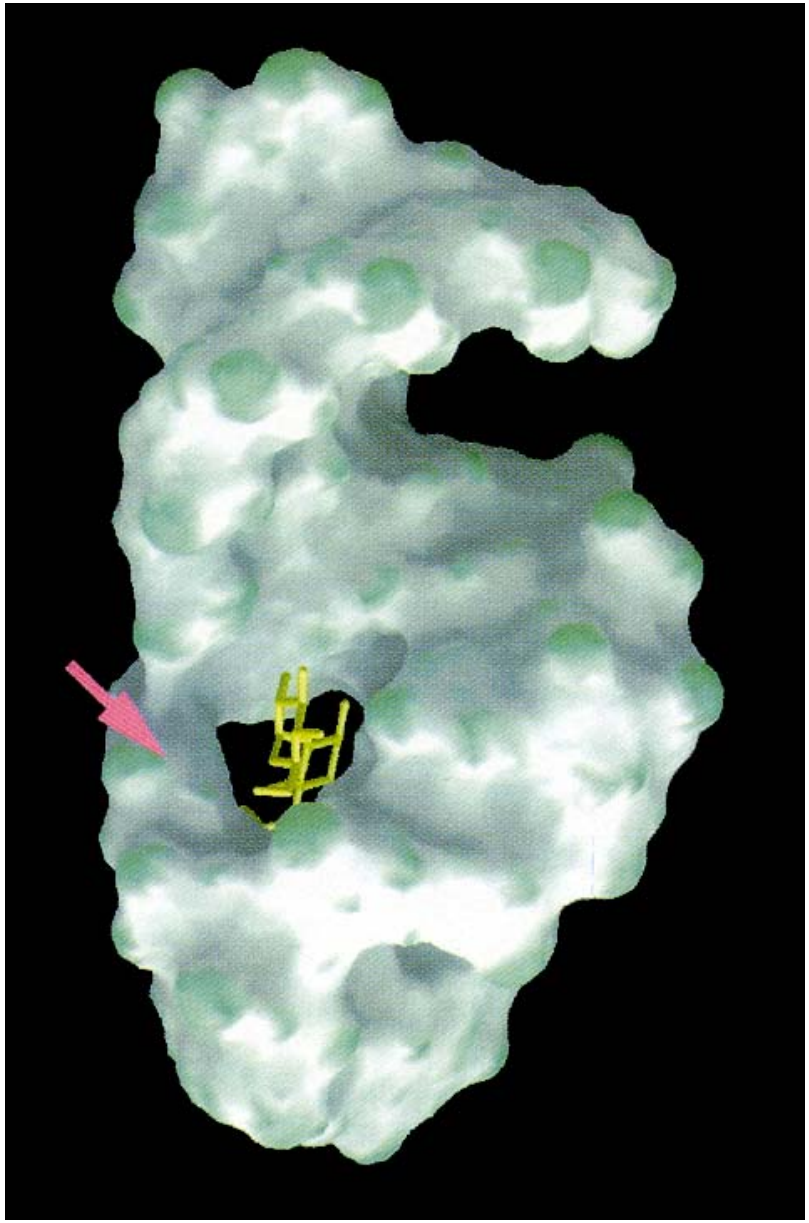
tobramycin



RNA aptamer, n = 27

Formation of secondary structure of the tobramycin binding RNA aptamer with **K_D = 9 nM**

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, ***Saccharide-RNA recognition in an aminoglycoside antibiotic-RNA aptamer complex.*** Chemistry & Biology **4**:35-50 (1997)



The three-dimensional structure of the tobramycin aptamer complex

L. Jiang, A. K. Suri, R. Fiala, D. J. Patel, Chemistry & Biology **4**:35-50 (1997)

Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer

ZHEN HUANG[1] and JACK W. SZOSTAK[2]

[1]Department of Chemistry, Brooklyn College, Ph.D. Programs of Chemistry and Biochemistry, The Graduate School of CUNY, Brooklyn, New York 11210, USA

[2]Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

ABSTRACT

Small changes in target specificity can sometimes be achieved, without changing aptamer structure, through mutation of a few bases. Larger changes in target geometry or chemistry may require more radical changes in an aptamer. In the latter case, it is unknown whether structural and functional solutions can still be found in the region of sequence space close to the original aptamer. To investigate these questions, we designed an in vitro selection experiment aimed at evolving specificity of an ATP aptamer. The ATP aptamer makes contacts with both the nucleobase and the sugar. We used an affinity matrix in which GTP was immobilized through the sugar, thus requiring extensive changes in or loss of sugar contact, as well as changes in recognition of the nucleobase. After just five rounds of selection, the pool was dominated by new aptamers falling into three major classes, each with secondary structures distinct from that of the ATP aptamer. The average sequence identity between the original aptamer and new aptamers is 76%. Most of the mutations appear to play roles either in disrupting the original secondary structure or in forming the new secondary structure or the new recognition loops. Our results show that there are novel structures that recognize a significantly different ligand in the region of sequence space close to the ATP aptamer. These examples of the emergence of novel functions and structures from an RNA molecule with a defined specificity and fold provide a new perspective on the evolutionary flexibility and adaptability of RNA.

Keywords: Aptamer; specificity; fold; selection; RNA evolution

RNA **9**:1456-1463, 2003

Evidence for neutral networks and shape space covering

Evidence for neutral networks and intersection of apatamer functions

J Mol Evol (2003) 57:299–308
DOI: 10.1007/s00239-003-2481-y

JOURNAL OF MOLECULAR EVOLUTION

Evolutionary Landscapes for the Acquisition of New Ligand Recognition by RNA Aptamers

Daniel M. Held, S. Travis Greathouse, Amit Agrawal, Donald H. Burke

Department of Chemistry, Indiana University, Bloomington, IN 47405-7102, USA

Received: 15 November 2002 / Accepted: 8 April 2003

Abstract. The evolution of ligand specificity underlies many important problems in biology, from the appearance of drug resistant pathogens to the re-engineering of substrate specificity in enzymes. In studying biomolecules, however, the contributions of macromolecular sequence to binding specificity can be obscured by other selection pressures critical to bioactivity. Evolution of ligand specificity *in vitro*—unconstrained by confounding biological factors—is addressed here using variants of three flavin-binding RNA aptamers. Mutagenized pools based on the three aptamers were combined and allowed to compete during *in vitro* selection for GMP-binding activity. The sequences of the resulting selection isolates were diverse, even though most were derived from the same flavin-binding parent. Individual GMP aptamers differed from the parental flavin aptamers by 7 to 26 mutations (20 to 57% overall change). Acquisition of GMP recognition coincided with the loss of FAD (flavin-adenine dinucleotide) recognition in all isolates, despite the absence of a counter-selection to remove FAD-binding RNAs. To examine more precisely the proximity of these two activities within a defined sequence space, the complete set of all intermediate sequences between an FAD-binding aptamer and a GMP-binding aptamer were synthesized and assayed for activity. For this set of sequences, we observe a portion of a neutral network for FAD-binding function separated from GMP-binding function by a distance of three mutations. Furthermore, enzymatic probing of these aptamers revealed gross structural remodeling of the RNA coincident with the switch in ligand recognition. The capacity for neutral drift along an FAD-binding network in such close approach to RNAs with GMP-binding activity illustrates the degree of phenotypic buffering available to a set of closely related RNA sequences—defined as the set's functional tolerance for point mutations—and supports neutral evolutionary theory by demonstrating the facility with which a new phenotype becomes accessible as that buffering threshold is crossed.

Key words: Aptamers — RNA structure — Phenotypic buffering — Fitness landscapes — Neutral evolutionary theory — Flavin — GMP

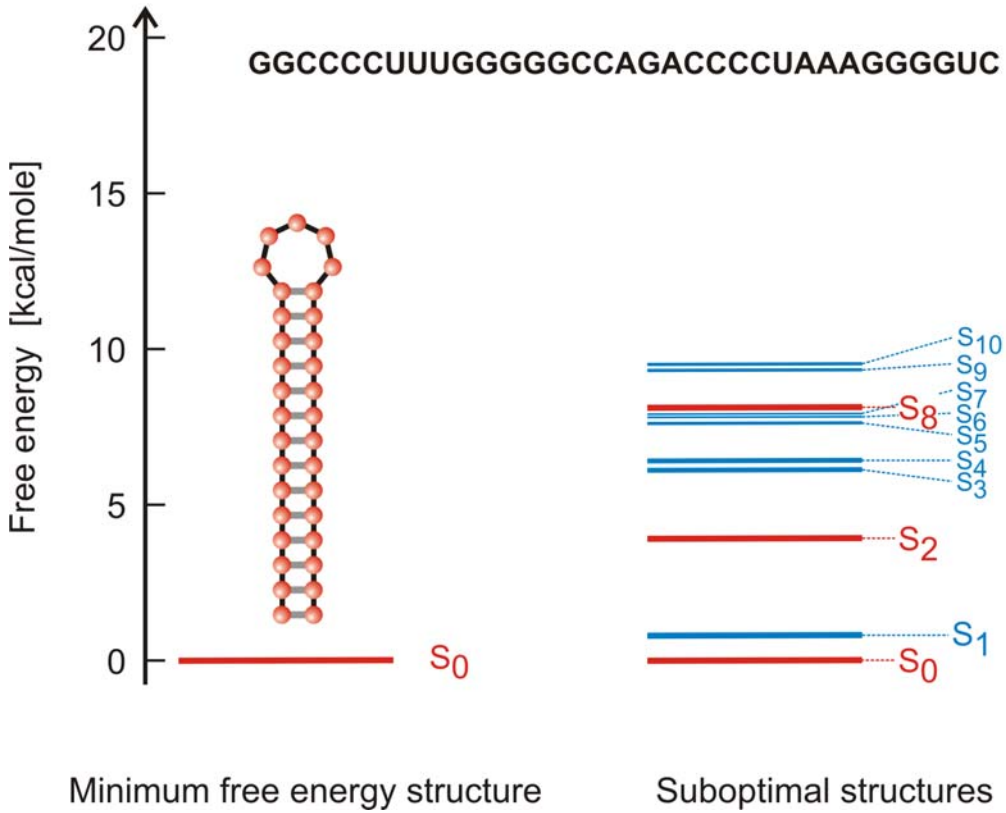
Correspondence to: Donald H. Burke; *email:* dhburke@indiana.edu

Introduction

RNA aptamers targeting small molecules serve as useful model systems for the study of the evolution and biophysics of macromolecular binding interactions. Because of their small sizes, the structures of several such complexes have been determined to atomic resolution by NMR spectrometry or X-ray crystallography (reviewed by Herman and Patel 2000). Moreover, aptamers can be subjected to mutational and evolutionary pressures for which survival is based entirely on ligand binding, without the complicating effects of simultaneous selection pressures for bioactivity, thus allowing the relative contributions of each activity to be evaluated separately.

The notion of RNA (secondary) structure

1. Minimum free energy structure
2. Many sequences one structure
3. **Suboptimal structures**
4. Kinetic structures



Extension of the notion of structure

Stefan Wuchty[1]
Walter Fontana[1,2]
Ivo L. Hofacker[1]
Peter Schuster[1,2]
[1] *Institut für Theoretische Chemie, Universität Wien, Währingerstrasse 17, A-1090 Wien, Austria*

[2] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501 USA*

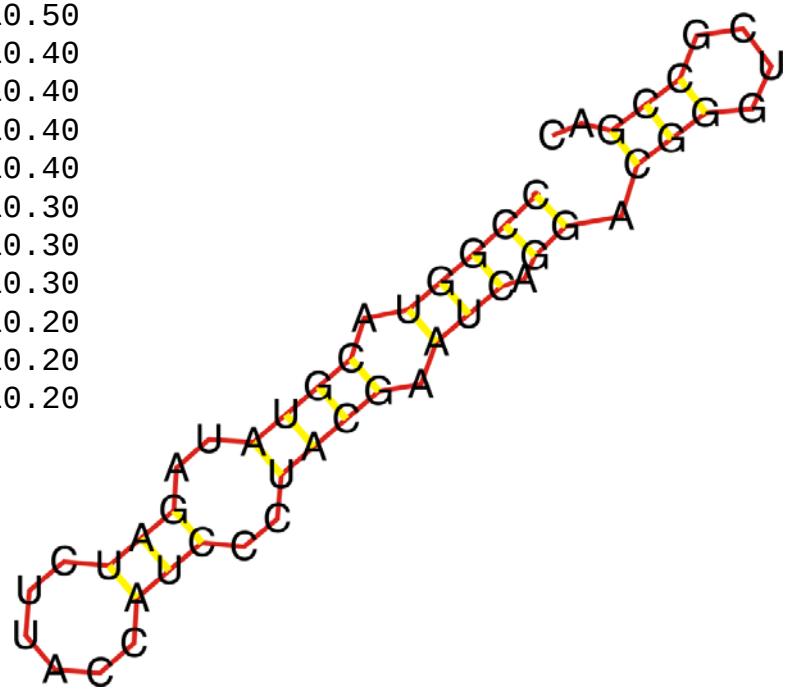
Received 13 May 1998; accepted 6 August 1998

Complete Suboptimal Folding of RNA and the Stability of Secondary Structures

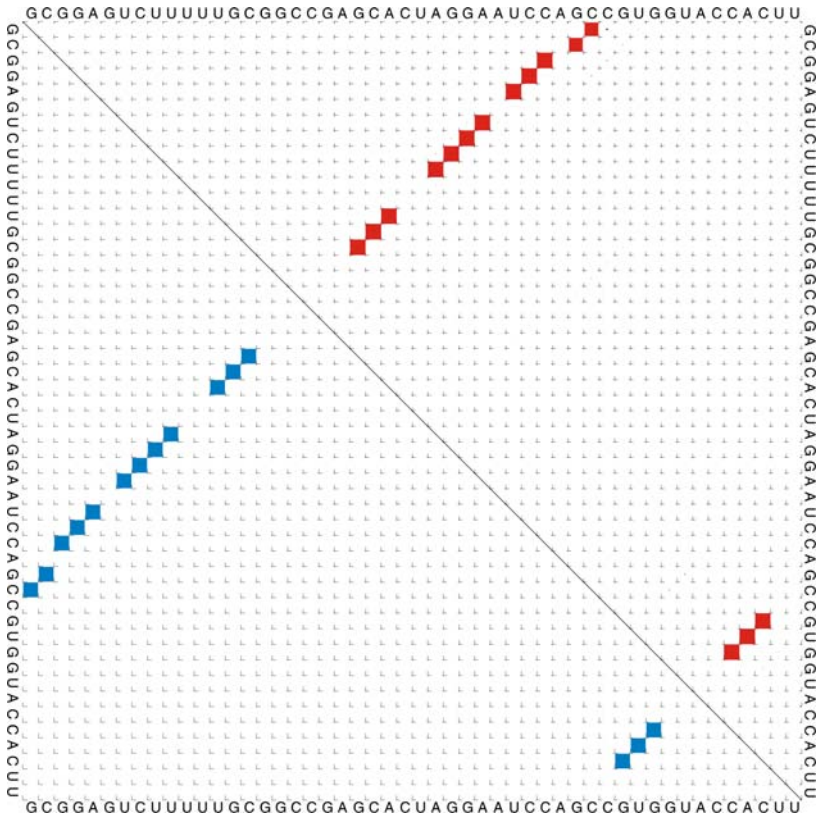
Abstract: *An algorithm is presented for generating rigorously all suboptimal secondary structures between the minimum free energy and an arbitrary upper limit. The algorithm is particularly fast in the vicinity of the minimum free energy. This enables the efficient approximation of statistical quantities, such as the partition function or measures for structural diversity. The density of states at low energies and its associated structures are crucial in assessing from a thermodynamic point of view how well-defined the ground state is. We demonstrate this by exploring the role of base modification in tRNA secondary structures, both at the level of individual sequences from* Escherichia coli *and by comparing artificially generated ensembles of modified and unmodified sequences with the same tRNA structure. The two major conclusions are that (1) base modification considerably sharpens the definition of the ground state structure by constraining energetically adjacent structures to be similar to the ground state, and (2) sequences whose ground state structure is thermodynamically well defined show a significant tendency to buffer single point mutations. This can have evolutionary implications, since selection pressure to improve the definition of ground states with biological function may result in increased neutrality.*
Biopoly 49: 145–165, 1999

Keywords: *RNA secondary structure; suboptimal folding; density of states; tRNA; modified bases; thermodynamic stability of structure; mutational buffering; neutrality; dynamic programming*

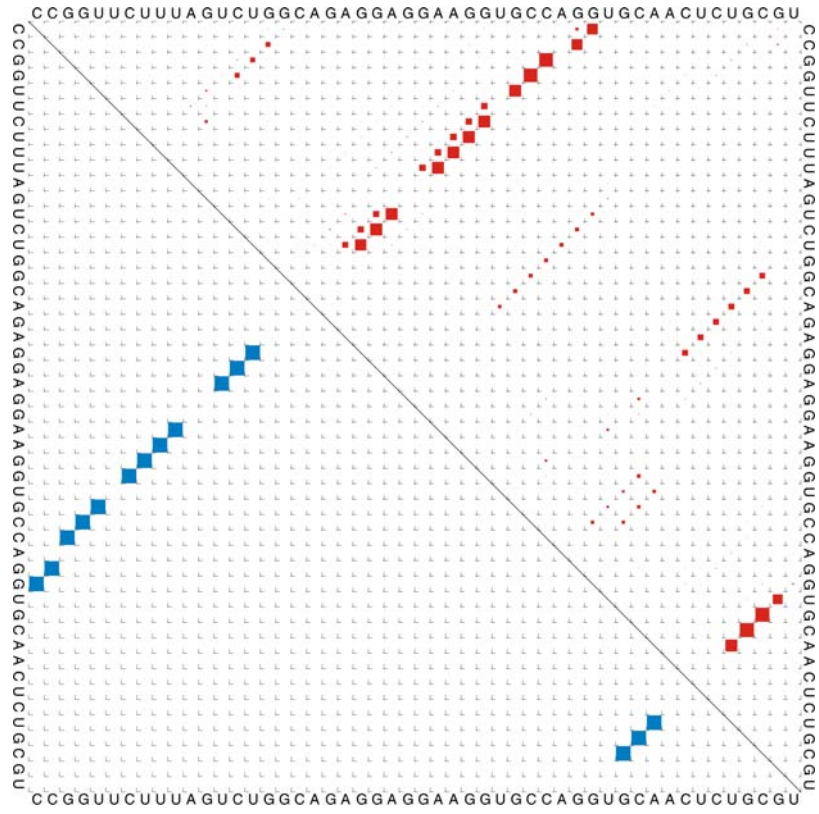
```
GCGUCGCGUGCCAUGGAGCAUCAUUACAUGAGACAGCCCCGGCCUCGGAU  -1220    200
(((((.((((..(((.....)))..)))).))).)).(((....))).. -12.20
(((((.((((..(((....)))..)))).))).)).(((....))).. -12.10
..(((.((.(((..((.((.((((...)))).....)))).)))).)))..)))) -11.50
..(((.((((..(((.....)))..)))).))).....(((....))).. -11.40
..(((.((((..(((....)))..)))).))).....(((....))).. -11.30
..(((.((.(((..((.((.(((.....))).....)))).)))).)))..)))) -11.30
..(((.((.(((..((.((.((((...))))....)).)))).)))..)))) -11.10
...(((.(.(((..((.((.((((...)))).....)))).)))).)))).. -11.10
..(((.((.(((..((.((.(((.....))).....)).)))).)))..)))) -10.90
...(((.(.(((..((.((.(((.....))).....)))).)))).)))).. -10.90
(((((.((((..(((.....)))..)))).))).)).((.....)).. -10.80
(((((.((((..(((....)))..)))).))).)).((.....)).. -10.70
...(((.(.(((..((.((.((((...)))).....)).)))).)))).. -10.70
..(((.((.(((..((....(((...)))).....))..)))..)))) -10.60
...((.((.(((..((.((.((((...)))).....)))).)))..)))). -10.60
...(((.(.(((..((.((.(((.....))).....)).)))).)))).. -10.50
....((.(.(((..((.((.((((...)))).....)))).)))).)).... -10.50
..(((.((((..(((.....)))..)))).))).((....)).....  -10.40
..(((.((.(((..((.((.(((.....))).....)))).)))).)))) -10.40
..(((.((.(((..((....(((.....)))....))..)))..)))) -10.40
...((.((.(((..((.((.(((.....))).....)))).)))..)))). -10.40
(((((.((((...((.....)).)))).))).)).(((....))).. -10.30
..(((.((((..(((....)))..)))).))).((....)).....  -10.30
....((.(.(((..((.((.(((.....))).....)))).)))).)).... -10.30
(((((.((((...((....)))...)))).))).)).(((....))).. -10.20
...(((.(.(((..((....(((...)))).....))..)))).))).. -10.20
...((.((.(((..((.((.((((...)))).....)).)))).)))). -10.20
              ..........................
              ..........................
              ..........................
```



GCGGAGUCUUUUUGCGGCCGAGCACUAGGAAUCCAGCCGUGGUACCACUU



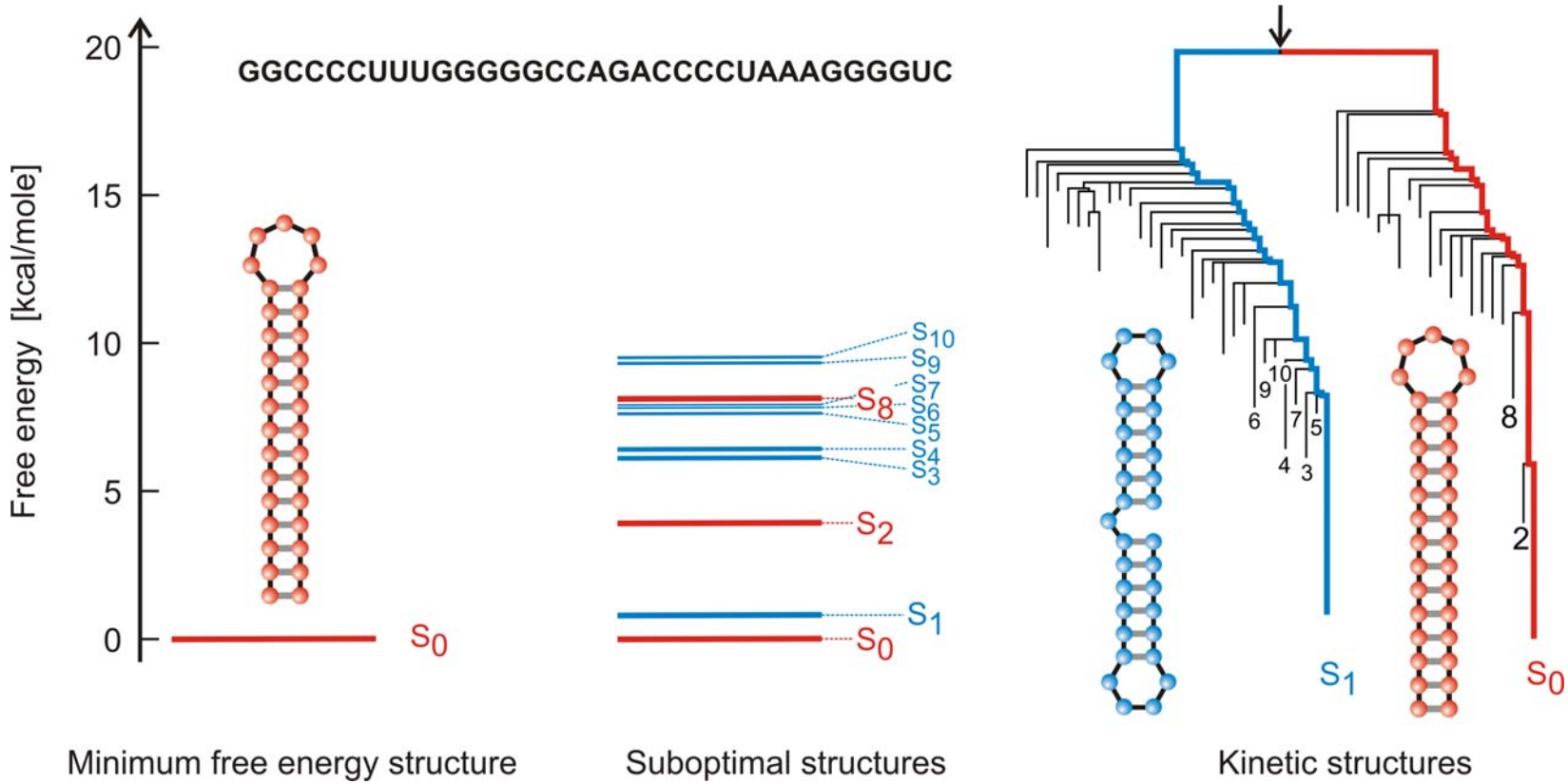
CCGGUUCUUUAGUCUGGCAGAGGAGGAAGGUGCCAGGUGCAACUCUGCGU



Two neutral sequences with very different contributions of suboptimal conformations

The notion of RNA (secondary) structure

1. Minimum free energy structure
2. Many sequences one structure
3. Suboptimal structures
4. **Kinetic structures**



Extension of the notion of structure

RNA (2000), *6*:325–338. Cambridge University Press. Printed in the USA.

RNA folding at elementary step resolution

CHRISTOPH FLAMM,[1] WALTER FONTANA,[2,3] IVO L. HOFACKER,[1] and PETER SCHUSTER[1]

[1]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, A-1090 Wien, Austria
[2]Santa Fe Institute, Santa Fe, New Mexico 87501 USA

ABSTRACT

We study the stochastic folding kinetics of RNA sequences into secondary structures with a new algorithm based on the formation, dissociation, and the shifting of individual base pairs. We discuss folding mechanisms and the correlation between the barrier structure of the conformational landscape and the folding kinetics for a number of examples based on artificial and natural sequences, including the influence of base modification in tRNAs.

Keywords: conformational spaces; foldability; RNA folding kinetics; RNA secondary structure

The Folding Algorithm

A sequence I specifies an energy ordered set of compatible structures $\mathfrak{S}(I)$:

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

A trajectory $\mathfrak{T}_k(I)$ is a time ordered series of structures in $\mathfrak{S}(I)$. 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:

$$\mathfrak{T}_0(I) = \{\mathbf{O}, S(1), \ldots, S(t-1), S(t), S(t+1), \ldots, \mathbf{S_0}\}$$
$$\mathfrak{T}_k(I) = \{\mathbf{O}, S(1), \ldots, S(t-1), S(t), S(t+1), \ldots, \mathbf{S_k}\}$$

Master equation

$$\frac{dP_k}{dt} = \sum_{i=0}^{m+1}\left(P_{ik}(t) - P_{ki}(t)\right) = \sum_{i=0}^{m+1} k_{ik} P_i - P_k \sum_{i=0}^{m+1} k_{ki}$$

$$k = 0,1,\ldots,m+1$$

Transition probabilities $P_{ij}(t) = \mathit{Prob}\{S_i \rightarrow S_j\}$ 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 dependent Ising models*. Phys.Rev. **145**:224-230, 1966).

Formulation of kinetic RNA folding as a stochastic process

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

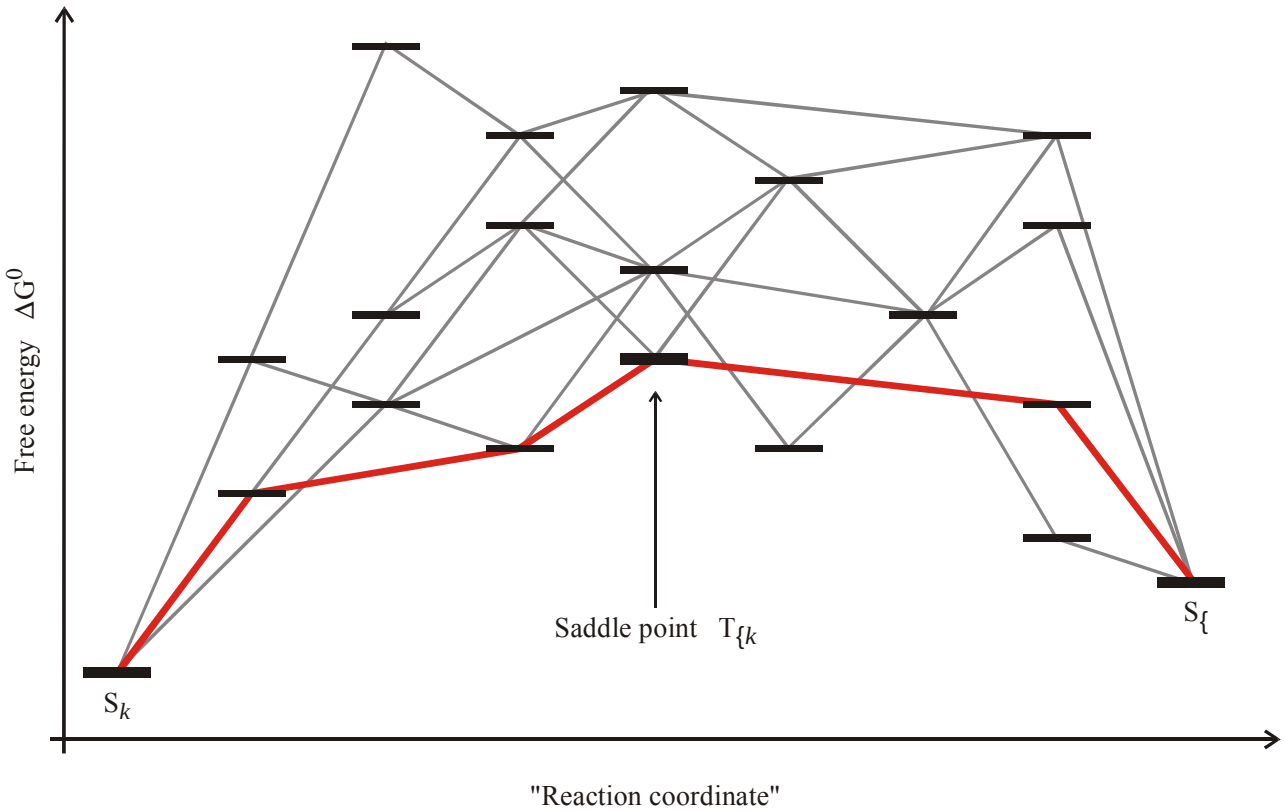
E-mail: ivo@tbi.univie.ac.at

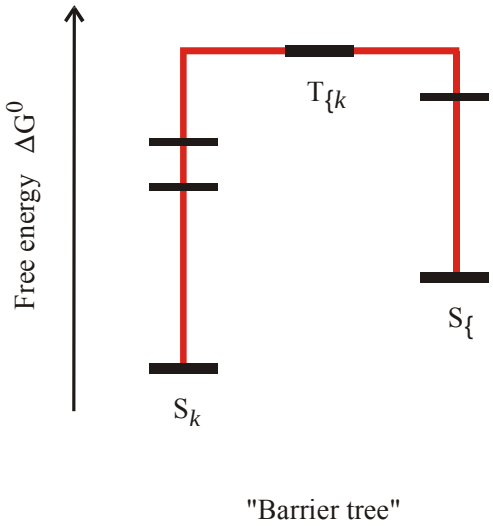
Received 22 December 2003, in final form 8 March 2004
Published 14 April 2004
Online at stacks.iop.org/JPhysA/37/4731 (DOI: 10.1088/0305-4470/37/17/005)

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





Definition of a ‚barrier tree‘

3568–3576 Nucleic Acids Research, 2006, Vol. 34, No. 12
doi:10.1093/nar/gkl445

Structural parameters affecting the kinetics of RNA hairpin formation

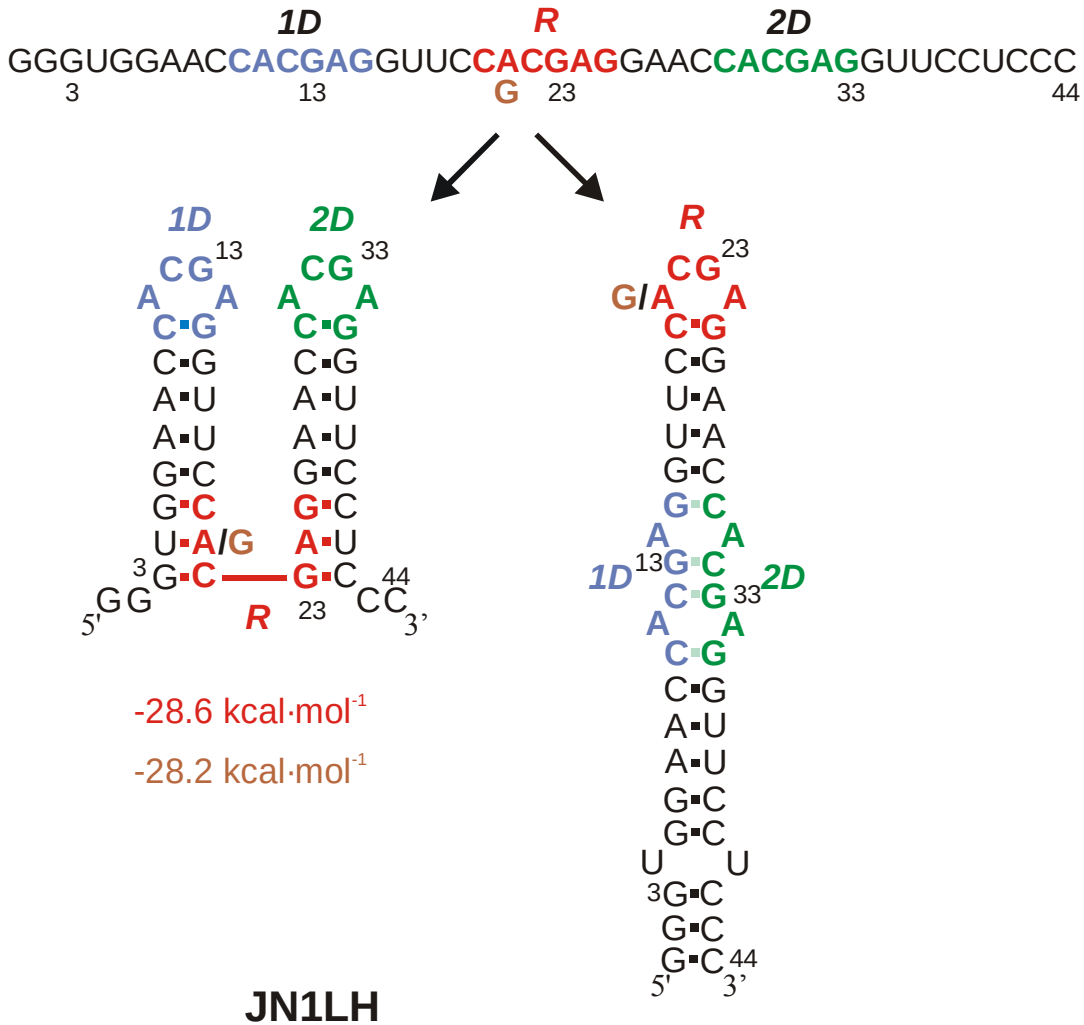
J. H. A. Nagel, C. Flamm[1], I. L. Hofacker[1], K. Franke[2], M. H. de Smit, P. Schuster[1] and C. W. A. Pleij*

Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, 2300 RA Leiden, The Netherlands, [1]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien, A-1090 Vienna, Austria and [2]IBA NAPS GmbH Rudolf-Wissell-Strasse 28 D-37079 Göttingen, Germany

Received January 28, 2005; Revised and Accepted June 7, 2006

ABSTRACT

There is little experimental knowledge on the sequence dependent rate of hairpin formation in RNA. We have therefore designed RNA sequences that can fold into either of two mutually exclusive hairpins and have determined the ratio of folding of the two conformations, using structure probing. This folding ratio reflects their respective folding rates. Changing one of the two loop sequences from a purine- to a pyrimidine-rich loop did increase its folding rate, which corresponds well with similar observations in DNA hairpins. However, neither changing one of the loops from a regular non-GNRA tetra-loop into a stable GNRA tetra-loop, nor increasing the loop size from 4 to 6 nt did affect the folding rate. The folding kinetics of these RNAs have also been simulated with the program 'Kinfold'. These simulations were in agreement with the experimental results if the additional stabilization energies for stable tetra-loops were not taken into account. Despite the high stability of the stable tetra-loops, they apparently do not affect folding kinetics of these RNA hairpins. These results show that it is possible to experimentally determine relative folding rates of hairpins and to use these data to improve the computer-assisted simulation of the folding kinetics of stem–loop structures.

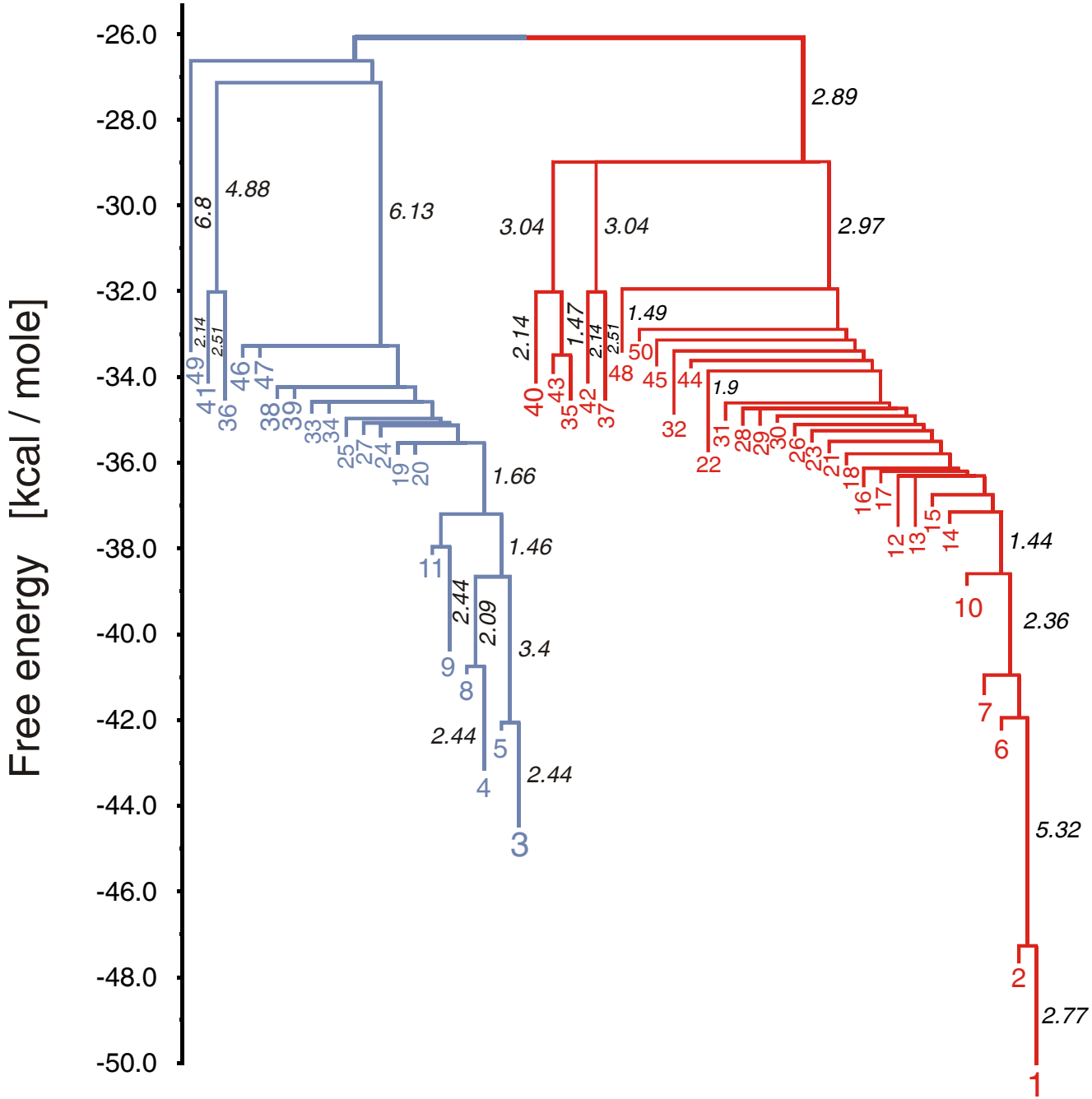


An experimental RNA switch

JN1LH

J.H.A. Nagel, C. Flamm, I.L. Hofacker, K. Franke, M.H. de Smit, P. Schuster, and C.W.A. Pleij.

Structural parameters affecting the kinetic competition of RNA hairpin formation. ***Nucleic Acids Res.*** **34**:3568-3576 (2006)



J1LH barrier tree

A ribozyme switch

E.A.Schultes, D.B.Bartel, Science **289** (2000), 448-452

minus the background levels observed in the HSP in the control (Sar1-GDP–containing) incubation that prevents COPII vesicle formation. In the microsome control, the level of p115-SNARE associations was less than 0.1%.
46. C. M. Carr, E. Grote, M. Munson, F. M. Hughson, P. J. Novick, *J. Cell Biol.* **146**, 333 (1999).
47. C. Ungermann, B. J. Nichols, H. R. Pelham, W. Wickner, *J. Cell Biol.* **140**, 61 (1998).
48. E. Grote and P. J. Novick, *Mol. Biol. Cell* **10**, 4149 (1999).
49. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
50. GST-SNARE proteins were expressed in bacteria and purified on glutathione-Sepharose beads using standard methods. Immobilized GST-SNARE protein (0.5 μM) was incubated with rat liver cytosol (20 mg) or purified recombinant p115 (0.5 μM) in 1 ml of NS buffer containing 1% BSA for 2 hours at 4°C with rotation. Beads were briefly spun (3000 rpm for 10 s) and sequentially washed three times with NS buffer and three times with NS buffer supplemented with 150 mM NaCl. Bound proteins were eluted three times in 50 μl of 50 mM tris-HCl (pH 8.5), 50 mM reduced glutathione, 150 mM NaCl, and 0.1% Triton X-100 for 15 min at 4°C with intermittent mixing, and elutes were pooled. Proteins were precipitated by MeOH/CH_3Cl and separated by SDS–polyacrylamide gel electrophoresis (PAGE) followed by immunoblotting using p115 mAb 13F12.
51. V. Rybin *et al.*, *Nature* **383**, 266 (1996).
52. K. G. Hardwick and H. R. Pelham, *J. Cell Biol.* **119**, 513 (1992).
53. A. P. Newman, M. E. Groesch, S. Ferro-Novick, *EMBO J.* **11**, 3609 (1992).
54. A. Spang and R. Schekman, *J. Cell Biol.* **143**, 589 (1998).
55. M. F. Rexach, M. Latterich, R. W. Schekman, *J. Cell Biol.* **126**, 1133 (1994).
56. A. Mayer and W. Wickner, *J. Cell Biol.* **136**, 307 (1997).
57. M. D. Turner, H. Plutner, W. E. Balch, *J. Biol. Chem.* **272**, 13479 (1997).
58. A. Price, D. Seals, W. Wickner, C. Ungermann, *J. Cell Biol.* **148**, 1231 (2000).
59. X. Cao and C. Barlowe, *J. Cell Biol.* **149**, 55 (2000).
60. G. G. Tall, H. Hama, D. B. DeWald, B. F. Horazdovsky, *Mol. Biol. Cell* **10**, 1873 (1999).
61. C. G. Burd, M. Peterson, C. R. Cowles, S. D. Emr, *Mol. Biol. Cell* **8**, 1089 (1997).
62. M. R. Peterson, C. G. Burd, S. D. Emr, *Curr. Biol.* **9**, 159 (1999).
63. M. G. Waters, D. O. Clary, J. E. Rothman, *J. Cell Biol.* **118**, 1015 (1992).
64. D. M. Walter, K. S. Paul, M. G. Waters, *J. Biol. Chem.* **273**, 29565 (1998).
65. N. Hui *et al.*, *Mol. Biol. Cell* **8**, 1777 (1997).
66. T. E. Kreis, *EMBO J.* **5**, 931 (1986).
67. H. Plutner, H. W. Davidson, J. Saraste, W. E. Balch, *J. Cell Biol.* **119**, 1097 (1992).
68. D. S. Nelson *et al.*, *J. Cell Biol.* **143**, 319 (1998).
69. We thank G. Waters for p115 cDNA and p115 mAbs; G. Warren for p97 and p47 antibodies; R. Scheller for rbet1, membrin, and sec22 cDNAs; H. Plutner for excellent technical assistance; and P. Tan for help during the initial phase of this work. Supported by NIH grants GM 33301 and GM42336 and National Cancer Institute grant CA58689 (W.E.B.), a NIH National Research Service Award (B.D.M.), and a Wellcome Trust International Traveling Fellowship (B.B.A.).

20 March 2000; accepted 22 May 2000

One Sequence, Two Ribozymes: Implications for the Emergence of New Ribozyme Folds

Erik A. Schultes and David P. Bartel*

We describe a single RNA sequence that can assume either of two ribozyme folds and catalyze the two respective reactions. The two ribozyme folds share no evolutionary history and are completely different, with no base pairs (and probably no hydrogen bonds) in common. Minor variants of this sequence are highly active for one or the other reaction, and can be accessed from prototype ribozymes through a series of neutral mutations. Thus, in the course of evolution, new RNA folds could arise from preexisting folds, without the need to carry inactive intermediate sequences. This raises the possibility that biological RNAs having no structural or functional similarity might share a common ancestry. Furthermore, functional and structural divergence might, in some cases, precede rather than follow gene duplication.

Related protein or RNA sequences with the same folded conformation can often perform very different biochemical functions, indicating that new biochemical functions can arise from preexisting folds. But what evolutionary mechanisms give rise to sequences with new macromolecular folds? When considering the origin of new folds, it is useful to picture, among all sequence possibilities, the distribution of sequences with a particular fold and function. This distribution can range very far in sequence space (*1*). For example, only seven nucleotides are strictly conserved among the group I self-splicing introns, yet secondary (and presumably tertiary) structure within the core of the ribozyme is preserved (*2*). Because these disparate isolates have the same fold and function, it is thought that they descended from a common ancestor through a series of mutational variants that were each functional. Hence, sequence heterogeneity among divergent isolates implies the existence of paths through sequence space that have allowed neutral drift from the ancestral sequence to each isolate. The set of all possible neutral paths composes a "neutral network," connecting in sequence space those widely dispersed sequences sharing a particular fold and activity, such that any sequence on the network can potentially access very distant sequences by neutral mutations (*3–5*).

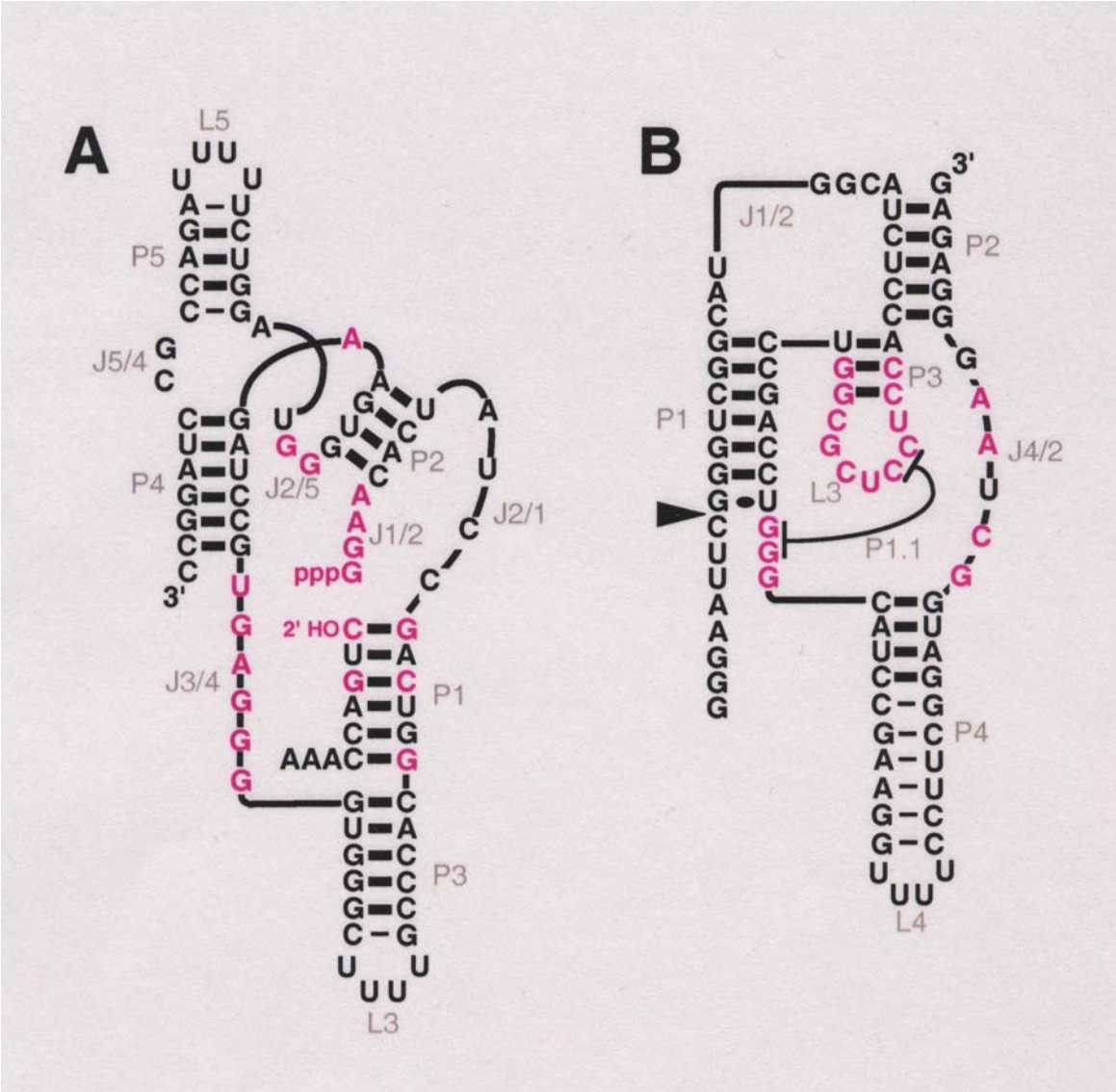
Theoretical analyses using algorithms for predicting RNA secondary structure have suggested that different neutral networks are interwoven and can approach each other very closely (*3*, *5–8*). Of particular interest is whether ribozyme neutral networks approach each other so closely that they intersect. If so, a single sequence would be capable of folding into two different conformations, would have two different catalytic activities, and could access by neutral drift every sequence on both networks. With intersecting networks, RNAs with novel structures and activities could arise from previously existing ribozymes, without the need to carry nonfunctional sequences as evolutionary intermediates. Here, we explore the proximity of neutral networks experimentally, at the level of RNA function. We describe a close apposition of the neutral networks for the hepatitis delta virus (HDV) self-cleaving ribozyme and the class III self-ligating ribozyme.

In choosing the two ribozymes for this investigation, an important criterion was that they share no evolutionary history that might confound the evolutionary interpretations of our results. Choosing at least one artificial ribozyme ensured independent evolutionary histories. The class III ligase is a synthetic ribozyme isolated previously from a pool of random RNA sequences (*9*). It joins an oligonucleotide substrate to its 5′ terminus. The prototype ligase sequence (Fig. 1A) is a shortened version of the most active class III variant isolated after 10 cycles of in vitro selection and evolution. This minimal construct retains the activity of the full-length isolate (*10*). The HDV ribozyme carries out the site-specific self-cleavage reactions needed during the life cycle of HDV, a satellite virus of hepatitis B with a circular, single-stranded RNA genome (*11*). The prototype HDV construct for our study (Fig. 1B) is a shortened version of the antigenomic HDV ribozyme (*12*), which undergoes self-cleavage at a rate similar to that reported for other antigenomic constructs (*13*, *14*).

The prototype class III and HDV ribozymes have no more than the 25% sequence identity expected by chance and no fortuitous structural similarities that might favor an intersection of their two neutral networks. Nevertheless, sequences can be designed that simultaneously satisfy the base-pairing requirements

Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

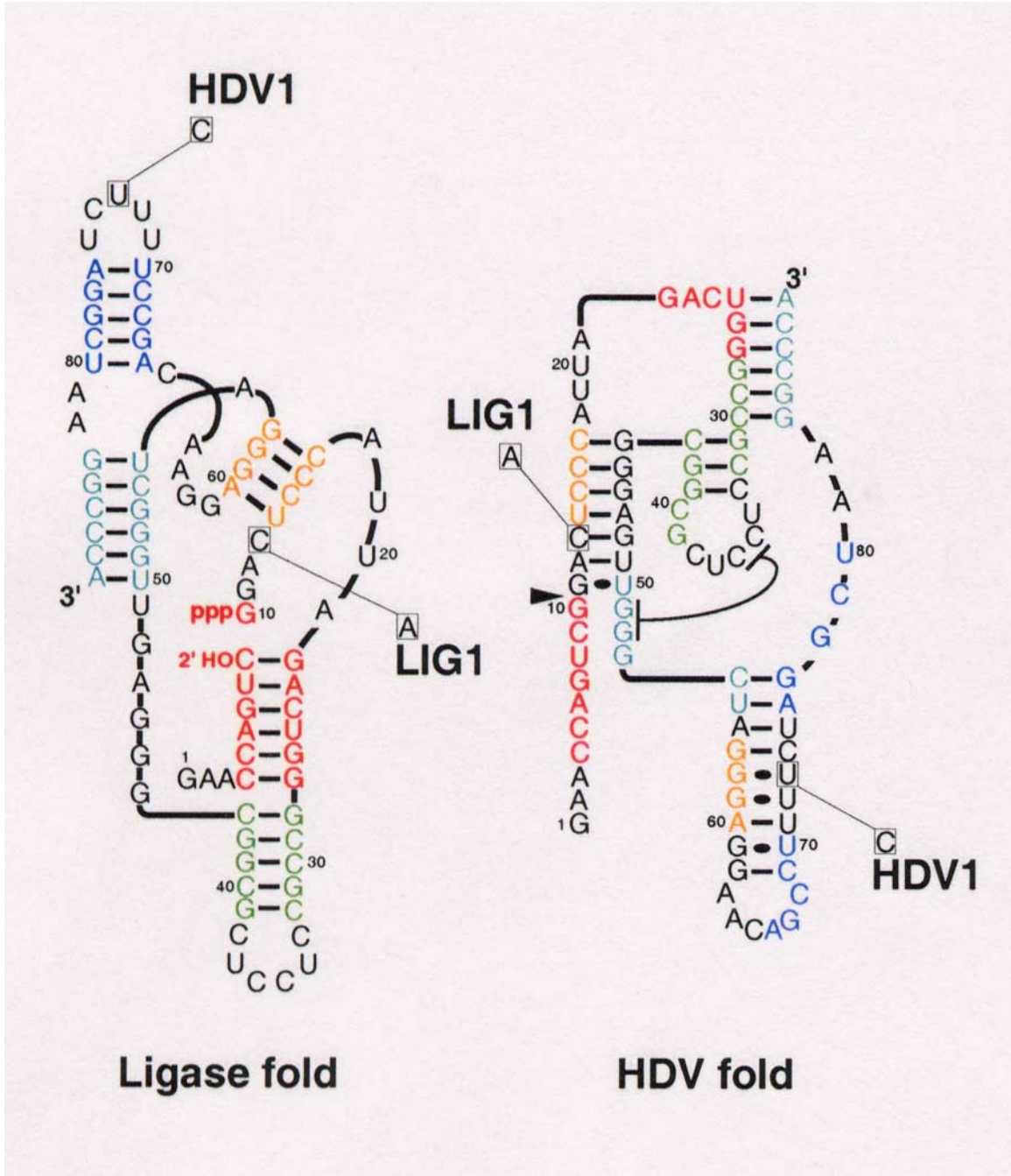
*To whom correspondence should be addressed. E-mail: dbartel@wi.mit.edu

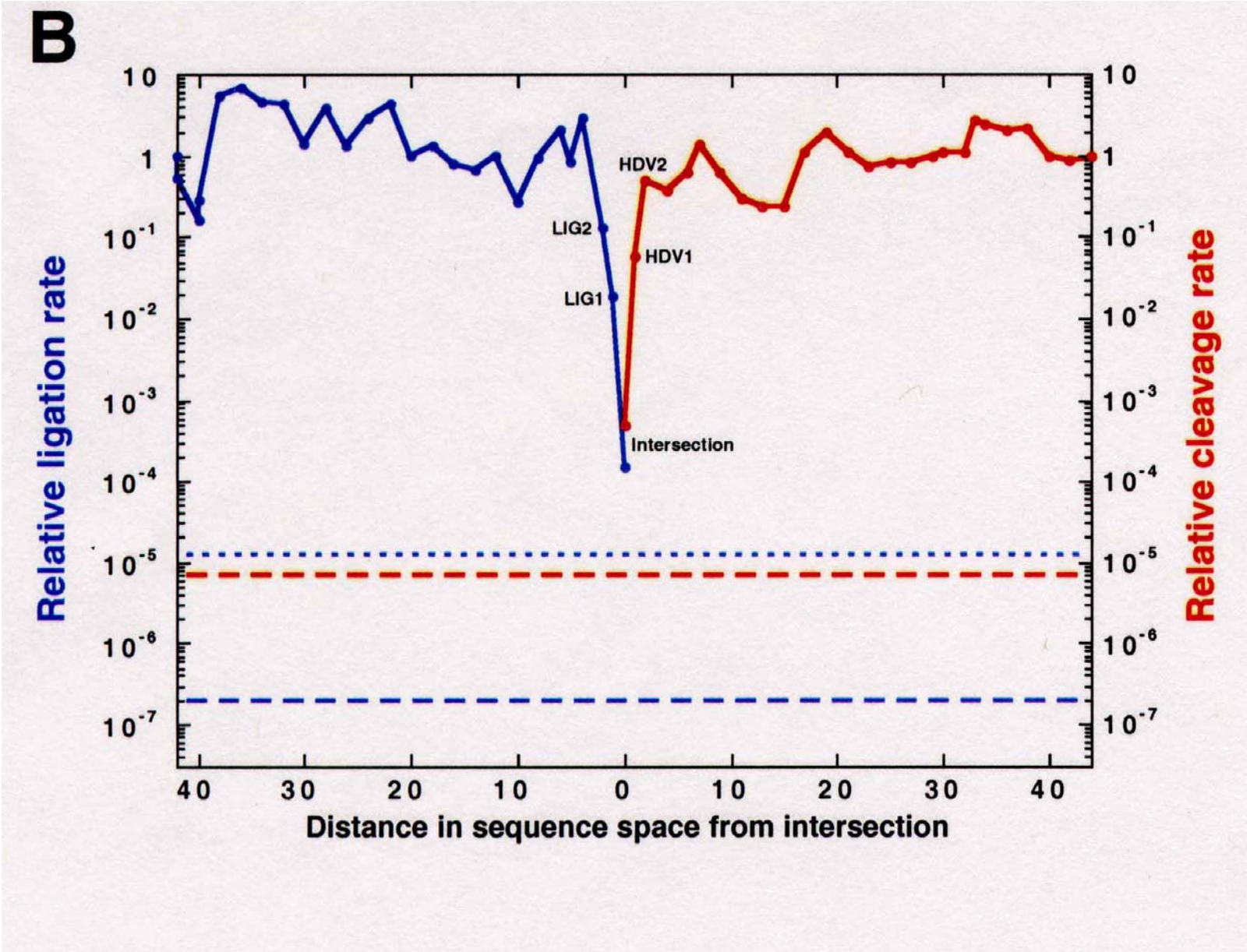


Two ribozymes of chain lengths n = 88 nucleotides: An artificial ligase (**A**) and a natural cleavage ribozyme of hepatitis-δ-virus (**B**)

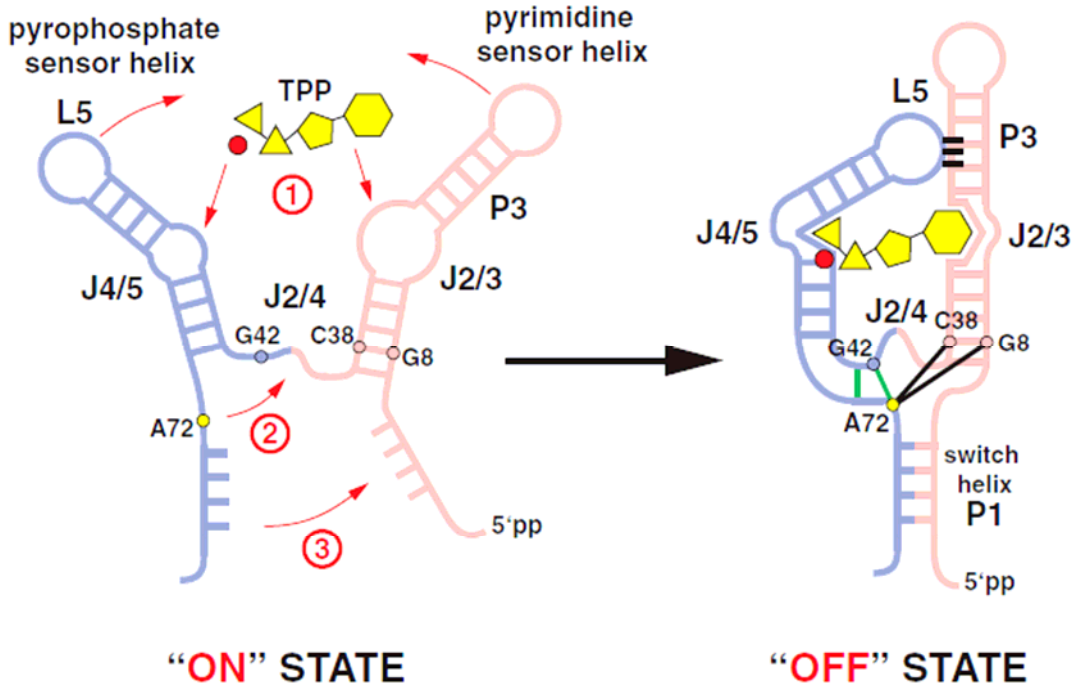
The sequence at the ***intersection***:

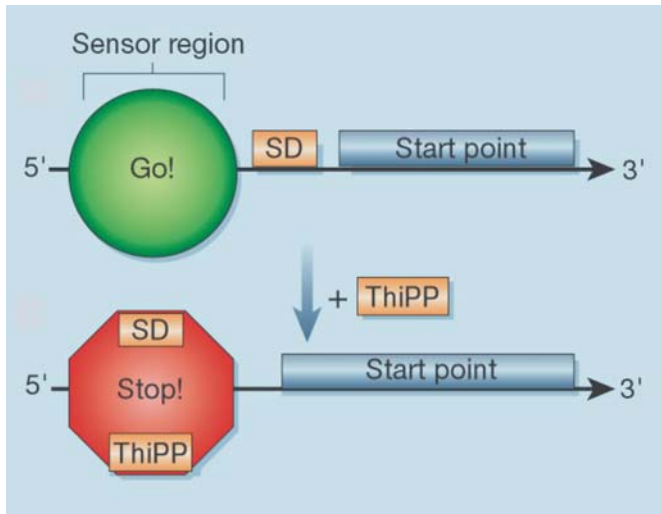
An RNA molecules which is 88 nucleotides long and can form both structures





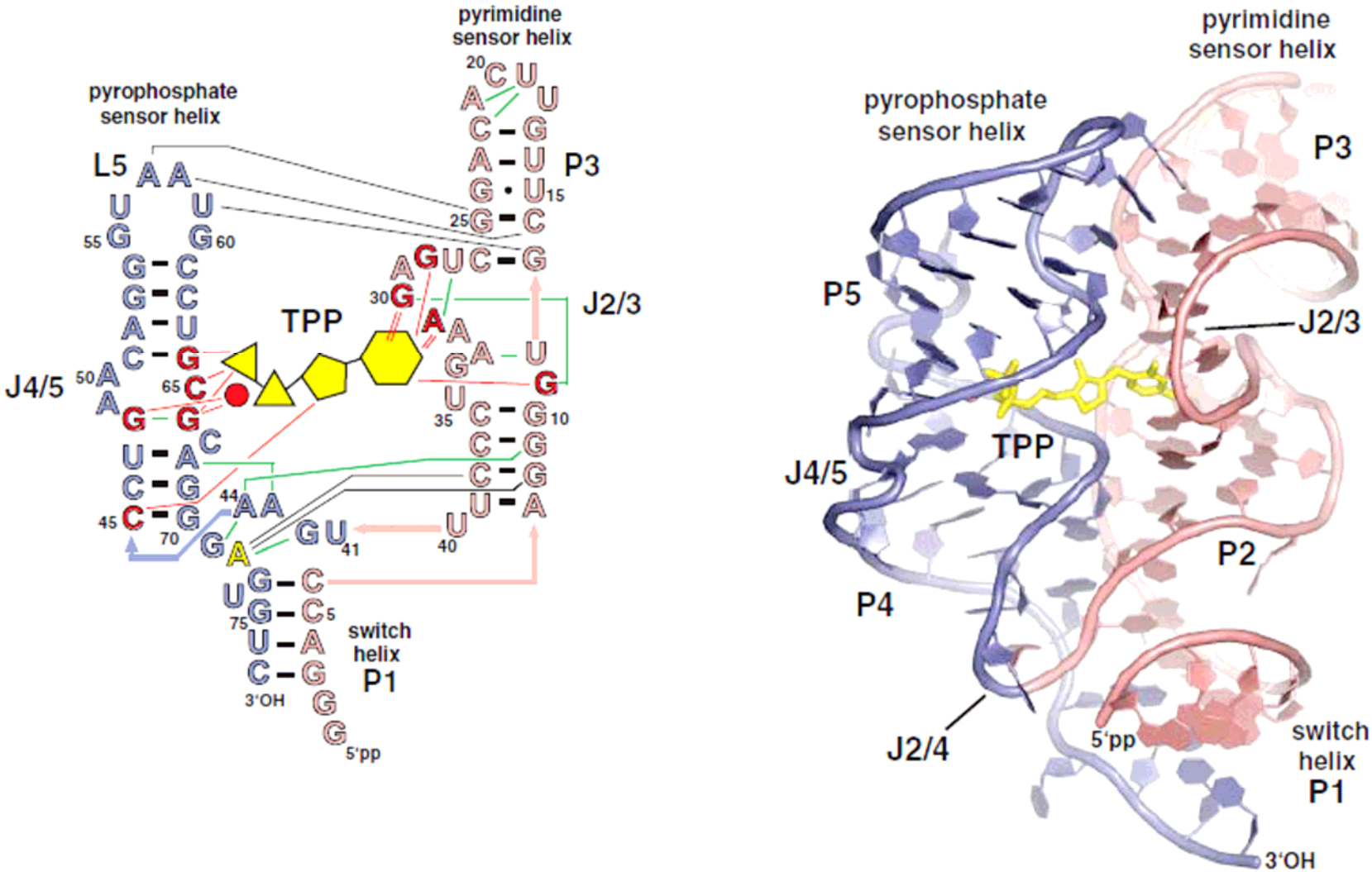
Two neutral walks through sequence space with conservation of structure and catalytic activity





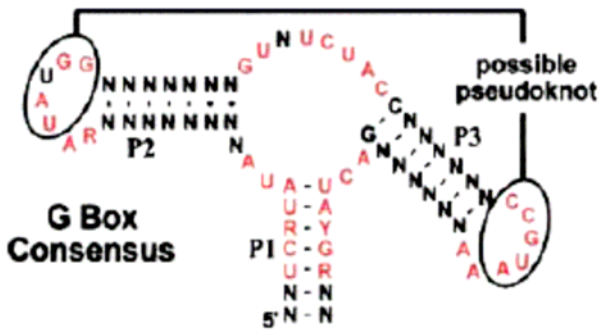
The thiamine-pyrophosphate riboswitch

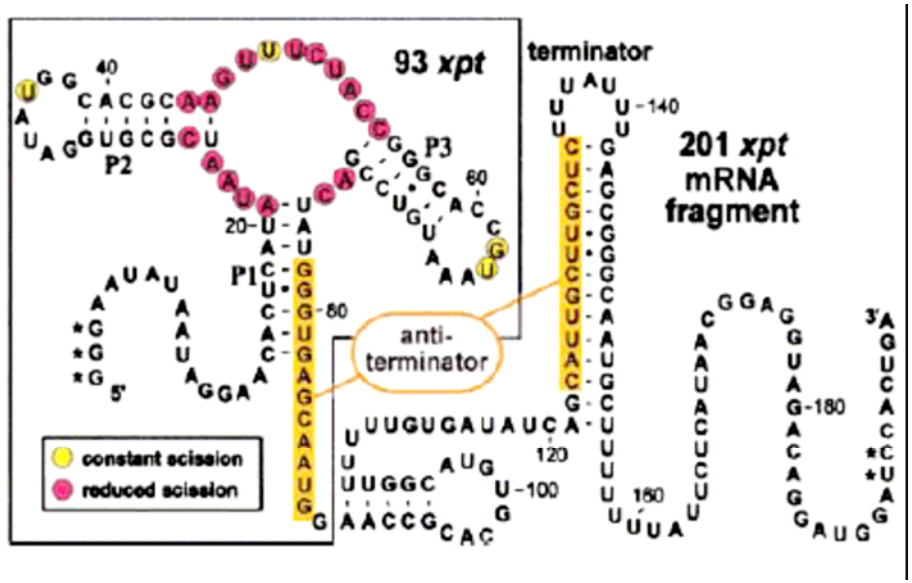
S. Thore, M. Leibundgut, N. Ban.
Science **312**:1208-1211, 2006.



The thiamine-pyrophosphate riboswitch

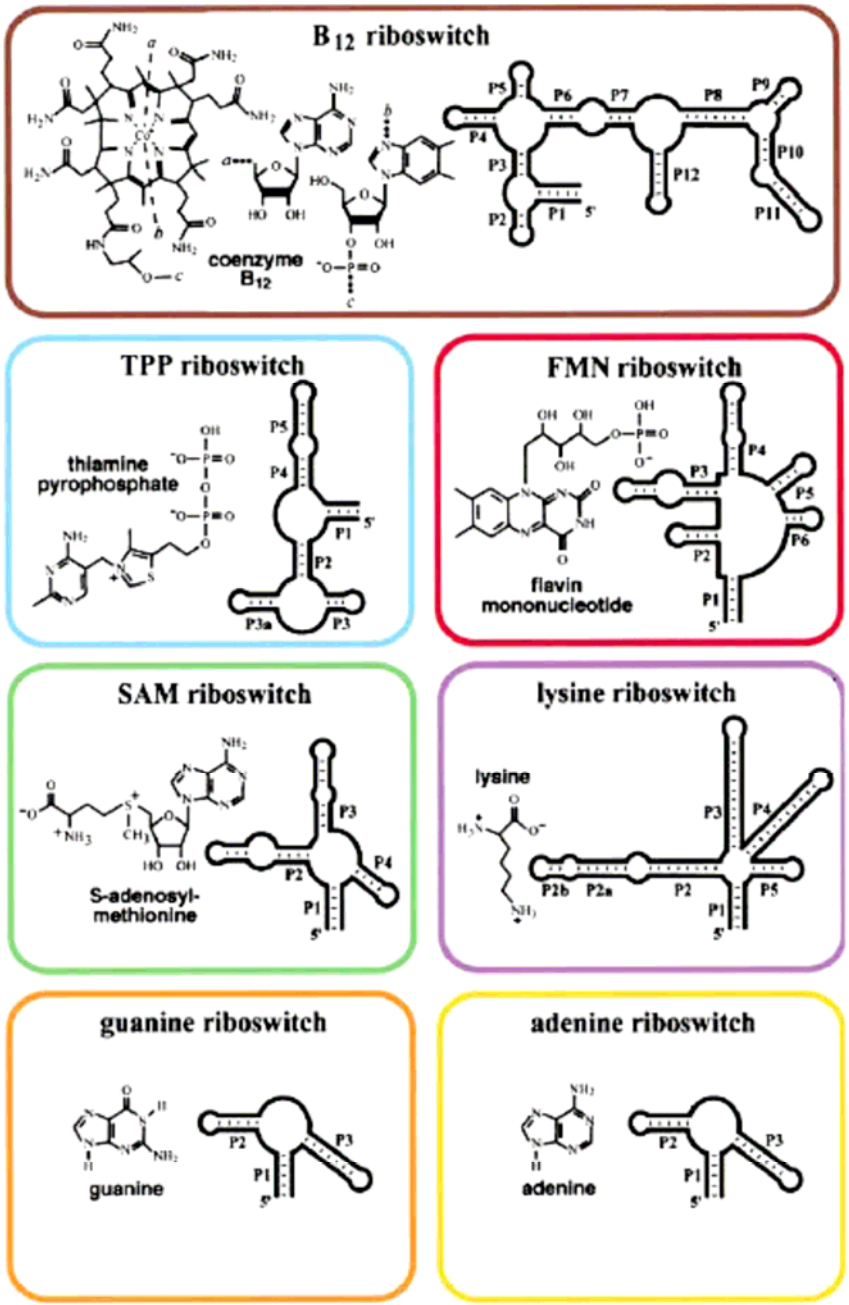
S. Thore, M. Leibundgut, N. Ban. *Science* **312**:1208-1211, 2006.

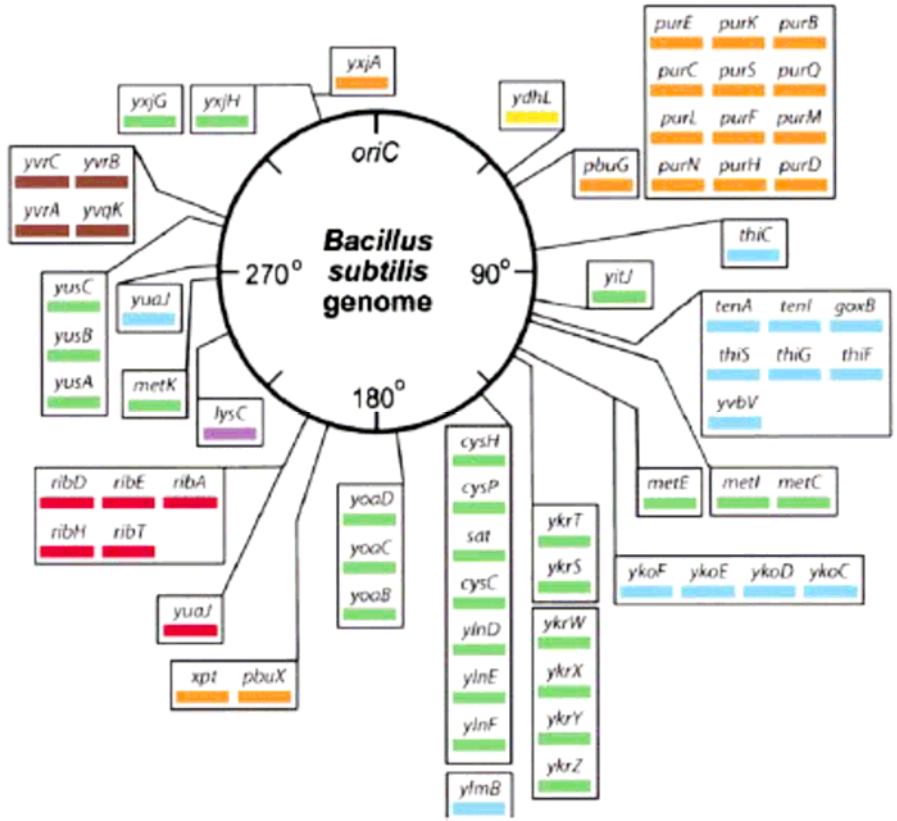




A natural metabolic riboswitch: The purine riboswitch

M. Mandal, B. Boese, J.E. Barrick, W.C. Winkler, R.R, Breaker, Cell 113:577-586 (2003)





M. Mandal, B. Boese, J.E. Barrick, W.C. Winkler, R.R, Breaker. Cell 113:577-586 (2003)

ENCODE stands for
ENCyclopedia Of DNA Elements.

ENCODE Project Consortium.
Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.
Nature **447**:799-816, 2007

Coworkers

Universität Wien

Peter Stadler, **Bärbel M. Stadler**, Universität Leipzig, GE

Paul E. Phillipson, University of Colorado at Boulder, CO

Heinz Engl, Philipp Kügler, **James Lu**, **Stefan Müller**, RICAM Linz, AT

Jord Nagel, **Kees Pleij**, Universiteit Leiden, NL

Walter Fontana, Harvard Medical School, MA

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

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

Ivo L.Hofacker, **Christoph Flamm**, **Andreas Svrček-Seiler**, Universität Wien, AT

Kurt Grünberger, Michael Kospach , **Andreas Wernitznig**, **Stefanie Widder, Stefan Wuchty**, **Andreas De Stefani,** Universität Wien, AT

Jan Cupal, **Stefan Bernhart**, **Lukas Endler, Ulrike Langhammer**, **Rainer Machne, Ulrike Mückstein**, **Hakim Tafer, Thomas Taylor,** Universität Wien, AT

Web-Page for further information:

http://www.tbi.univie.ac.at/~pks

Review article:

Peter Schuster. Prediction of RNA molecules: From theory to models and real molecules. *Rep.Prog.Phys.* **69**:1419-1477, 2006.


Happy Birthday, dear Janos!