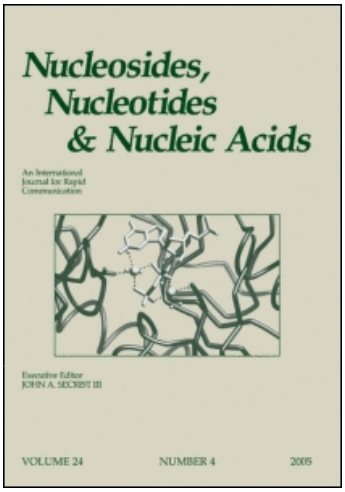


Nucleosides, Nucleotides and Nucleic Acids
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

A CODE RELATING SEQUENCE TO STRUCTURE IN NUCLEIC ACIDS

Mieczysław Remin[a]
[a] Division of Biophysics, Institute of Experimental Physics, Department of Physics, University of Warsaw, Warszawa, Poland

Online publication date: 31 March 2001

To cite this Article Remin, Mieczysław(2001) 'A CODE RELATING SEQUENCE TO STRUCTURE IN NUCLEIC ACIDS', Nucleosides, Nucleotides and Nucleic Acids, 20: 4, 841 — 846
To link to this Article: DOI: 10.1081/NCN-100002442
URL: http://dx.doi.org/10.1081/NCN-100002442

NUCLEOSIDES, NUCLEOTIDES & NUCLEIC ACIDS, 20(4–7), 841–846 (2001)

A CODE RELATING SEQUENCE TO STRUCTURE IN NUCLEIC ACIDS

Mieczysław Remin

Division of Biophysics, Institute of Experimental Physics,
Department of Physics, University of Warsaw, al. Żwirki
i Wigury 93, Warszawa, PL 02-089, Poland

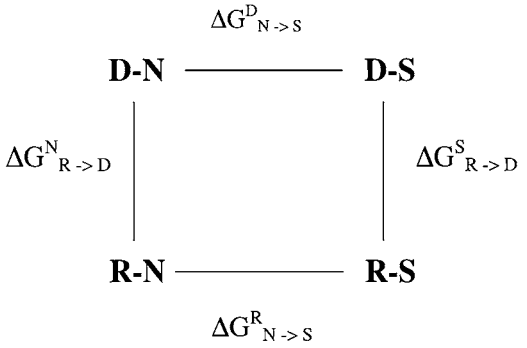
ABSTRACT

Nucleic acids are elucidated in configuration space. An algorithm relating sequence to stability in A and B helical secondary structures, is stated to incorporate NMR conformational and optical melting data. This made possible a classification of elementary sequences in terms of configuration forces driving between A and B states, a finding useful in prediction of structural behavior of different sequences of DNA, RNA and their hybrids.

Molecular evolution has used two forms of the sugar link in the backbone of nucleic acids, ribose and deoxyribose. This attribute is responsible for fundamental differences in structural behavior of DNA and RNA, eventually determining their biological functions. This work insights into several important questions. In particular, what is the role of sequence and functional groups, especially the sugar ring and nucleic bases, in driving of RNA and DNA to their stable forms in aqueous solution? Why are RNA duplexes usually more stable than corresponding DNA and why does their relative stability depends on sequence? What are general rules relating configuration states of RNA and DNA and how do they determine properties of their hybrids? How do configuration states determine conditions of conformational transition, as a criterion in classification of sequences? How do these questions provide insight into the path of evolution of the genetic code?

*Corresponding author.

SCHEME AND ALGORITHM CONFIGURATION FORCES CONSTITUTING THERMODYNAMIC CYCLE



SYMBOL	EXPRESSION	PHYSICAL INTERPRETATION
$\Delta G^{R}_{N \to S}$	$-RT\ln(P_{RS}/P_{RN})$	Free energy associated with N → S conformational change
$\Delta G^{D}_{N \to S}$	$-RT\ln(P_{DS}/P_{DN})$	in ribo- and deoxy- nucleotides P_{ij} – sugar pucker populations
$F_S(x)$	$\Delta G^{D}_{N \to S} - \Delta G^{R}_{N \to S}$	Driving force of the sugar cycle: 2'H replacement of 2'OH effect
$F_F(x)$	$\frac{1}{2}(\Delta G^{D}_{N \to S} + \Delta G^{R}_{N \to S})$	N→ S driving of functional groups: x = A, G, U/T and C
$\Delta G^{R}_{N \to S}$	$-\frac{1}{2}F_S(x) + F_F(x)$	Configuration forces of the N→ S free energy differences
$\Delta G^{D}_{N \to S}$	$\frac{1}{2}F_S(x) + F_F(x)$	in RNA and DNA constituents

Free Energy Approach

A complex transformation of a molecular system can be considered as a thermodynamic process consisting of several steps in a multidimensional configuration space. The steps may represent reactions of a physical or chemical nature. A reaction might be completed in a microseconds or take thousands of years. The Gibbs free energy (G), as the function of state, incorporates the requirements of both the First and Second Laws, and provides a criterion for spontaneity of the reaction. This approach follows an idea of a thermodynamic cycle for changes of state of the sugar ring (1). The idea uses the principle, originally known as Hess's Law, stating that a state function for the entire process is independent of path (2).

Table 1. NMR Couplings (Hz), Conformational Populations (%) and Configuration Forces (kcal/mol) for 3′-Phosphate Nucleotides[a]

Symbol	$J_{3'4'}$	P_{RN}	P_{RS}	$J_{3'4'}$	P_{DN}	P_{DS}	$F_S(x)$	$F_F(x)$	$\frac{1}{2}F_S + F_F$	$-\frac{1}{2}F_S + F_F$
A	3.2	27.3	72.7	2.5	19.2	80.8	−0.266	−0.704	−0.837	−0.571
G	3.6	32.5	67.5	3.0	26.0	74.0	−0.182	−0.517	−0.608	−0.426
U*	5.4	55.8	44.2	3.2	28.8	71.2	−0.665	−0.196	−0.528	0.137
C	5.8	61.0	39.0	3.5	32.9	67.1	−0.677	−0.077	−0.416	0.261

[a]A, G, U*, C, denote type of the base in nucleotide, where U* means uracil in ribo-type and thymine in deoxy-type; The molar fractions in aqueous solution: P_{RN}, P_{RS}, P_{DN}, P_{DS}, [%] of N (C3′*endo*) and S (C2′*endo*) conformational ensembles of the sugar ring, of ribo-type (R) or deoxy-type (D), are determined from vicinal proton-proton coupling constants, $J_{3'4'}$ [in Hz] (data source in ref. 5–6); according to procedure presented in ref. 2, 8–10. 3′-AMP, 3′-GMP, 3′-UMP were measured as their sodium salts and 3′-CMP as the lithium salt (5–6). Configuration forces: $F_S(x) = \Delta G^D_{N\to S} - \Delta G^R_{N\to S}$, is the driving force of the sugar cycle, and the module $F_F(x) = \frac{1}{2}(\Delta G^D_{N\to S} + \Delta G^R_{N\to S})$, represents the N → S driving of functional groups, where free energy differences (in kcal/mol nucleotide), $\Delta G^R_{N\to S} = -RT\ln(P_{RS}/P_{RN})$, and $\Delta G^D_{N\to S} = -RT\ln(P_{DS}/P_{DN})$, (see Algorithm).

Configuration Space

This work introduces a model of three-dimensional configuration space where each of the dimensions distinguishes two basic coordinates representing discrete states. The dimensions and coordinates are: ***Chemical Structure***, where RNA and DNA are related to the two forms of the sugar link in the backbone of nucleic acids, ribose and deoxyribose; ***Molecular Conformation***, where A and B canonical forms represent two principal families of helical secondary structures (3), and are generally related to two puckering domains of the sugar ring, *N-C3′ endo and S-C2′ endo*; ***Degree of Order***, where the coil phase is free of base stacking and base pairing, the interactions present in the double helix.

Thermodynamic Cycles

A sequence determines its eight states, each characterised by three coordinates, one of the two in each dimension. Configuration states constitute two types of cycles in this space. A "***Sugar Cycle***" relates changes of molecular conformation to changes of chemical structure, and is dependent on composition in the coil phase. A "***Phase Cycle***" of RNA and DNA, relates changes of molecular conformation to changes of degree of order, and is dependent on sequence. They are determined by incorporation of NMR conformational (sugar cycle) and free energy of duplex formation data from optical melting (phase cycle).

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Conformational Characteristics

They are essentially different for RNA and DNA constituents (4). As representatives of nucleotides in the coil phase, they provide reference conformations for two different helical secondary structures, A-RNA and B-DNA. Nucleic bases also play an important role in conformational characteristics of the sugar ring in RNA and DNA constituents (4). As determined here, in opposition to pyrimidines, purines generate a significantly greater $N\,(C3'endo) \rightarrow S\,(C2'endo)$ driving, a force important in the structure of nucleic acids.

Rules Relating Stability of DNA, RNA and their Hybrids

The hybrid states are assumed as a superposition of DNA and RNA states. This way stability of model hybrids in A and B helical forms, is unequivocally determined by stability of DNA and RNA. This determines several rules. For example, sum of DNA and RNA stability vs. sum of model hybrids stability, is constant and equal half of the "driving force" of the sugar cycle. The rules generally well predict stability of real hybrids. Experimental verification is performed for a variety of sequences of 8–21 base pairs long of DNA, RNA and their hybrids, for a particular set of conditions of optical melting (5).

Classification of Elementary Sequences

The idea uses the nearest-neighbor approach assuming that each sequence of DNA or RNA, can be considered as a composition of 10 possible elementary sequences (6–10). Incorporation of empirical nearest-neighbor thermodynamic parameters from optical melting study, made possible a determination of configuration forces driving particular elementary sequences between A and B helical forms.

Finally all elementary seque nces have been classified in terms of the criterion of configuration forces driving between A and B helical forms. They are in the following order: **GG/CC**, **GC/CG**, **CU*/GA**, **GA/CU***, **GU*/CA**, **CA/GU***, **U*A/AU***, **CG/GC**, **AU*/U*A** and **AA/U*U***, where U* means uracil in RNA and thymine in DNA The classification is universal thus is useful in prediction of structural behavior of different sequences of DNA, RNA and their hybrids.

The first and the last sequence of this order, with extreme values of configuration forces, show opposite tendencies: **GG/CC** in DNA, reveals a significant, but in RNA a very strong, B → A driving; in hybrids, A-form is significantly more stable than B-form. **AA/U*U*** in DNA, reveals a strong A → B driving, whereas in RNA, the both A and B forms show similar stability; in hybrids, B-form is considerable more stable than A-form.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Two others, GC/CG and CG/GC, of the same composition, show very different behavior: **GC**/**CG** in DNA reveals a moderate, but in RNA a strong, B → A driving; in hybrids, A-form is considerable more stable than B-form. **CG**/**GC** in DNA reveals a significant A → B driving, whereas in RNA a considerable B → A driving; in hybrids, the both forms, A and B, have similar stability, with a slight preference of B-form.

Evolutionary Implications

The findings explain why are RNA duplexes usually more stable than corresponding DNA and why does their relative stability depends on sequence. Secondary structure forces, including sequence dependent base stacking, are found here to be significantly greater in A-form than in B-form. This indicates on a fundamental role of RNA in an early stage of molecular evolution of nucleic acids and the genetic code based on DNA. The findings uncover the nature of sequence-dependent deformability of DNA, an important determinant of its ability to interact with proteins. Due to this, DNA is not only a carrier of genetic information but plays an important role in the processes of enzymatic recognition.

ACKNOWLEDGMENTS

This work was supported by the KBN Program BST661/BF.

REFERENCES

1. Remin, M. Thermodynamic cycle between DNA and RNA constituents for conformation of the sugar ring from nuclear magnetic resonance study. *J. Biomol. Struct. & Dynam.* **1997**, *15*, 251–64.
2. Freifelder, D. *Principles of Physical Chemistry with Applications to the Biological Sciences, Jones and Barlett Publishers, Inc., Boston* **1985**.
3. Saenger, W. in *Principles of Nucleic Acid Structure. Springer-Verlag, N.Y.* **1984**.
4. Remin, M. The nature of conformational correlations in α- and β-anomers of nucleic acid components by nuclear magnetic resonance. *J. Biomol. Struct. & Dynam.* **1988**, *6*, 367–382.
5. Lesnik, E. A.; Freier, S. M. Relative thermodynamic stability of DNA, RNA and their hybrid duplexes: relationship with base composition and structure. *Biochemistry* **1995**, *34*, 10807–15.
6. Turner, D. H. Thermodynamics of base pairing. *Curr. Opinion in Struct. Biol.* **1996**, *6*, 299–304.
7. SantaLucia, J., Jr. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **1996**, *35*, 3555–3562.
8. SantaLucia, J., Jr. and Turner, D. H. Measuring the thermodynamics of RNA secondary structure formation. *John Wley and Sons, Inc.* **1998**, 309–319.

9. SantaLucia, J., Jr. A united view of polymer, dumbbell and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1460–1465.
10. Xia, T. B. et al. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **1998**, *37*, 14719–35.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Request Permission or Order Reprints Instantly!