

Academy of Sciences of the Czech Republic
Institute of Organic Chemistry and Biochemistry, v.v.i.
Department of Molecular Spectroscopy

and

Charles University in Prague
Faculty of Mathematics and Physics



Ph.D. Thesis

**Calculation of the NMR spectroscopy parameters and
their structural interpretation in biomolecules**

RNDr. Zuzana Vokáčová

Supervizor: Dr. Vladimír Sychrovský

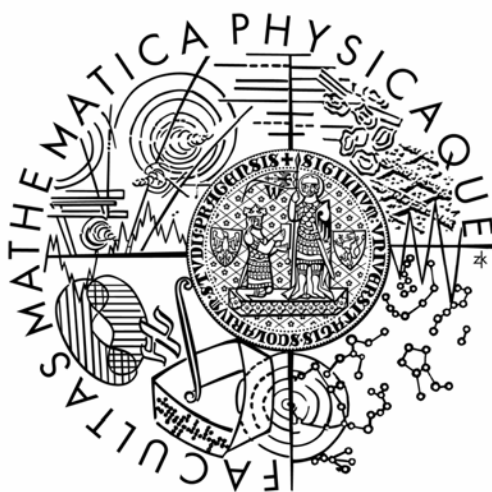
Consultants: Doc. RNDr. Ing. Jaroslav Burda, CSc., Prof. RNDr. Jiří Šponer, DrSc.

Prague 2009

Akademie věd České republiky
Ústav organické chemie a biochemie, v.v.i.
Oddělení molekulární spektroskopie

a

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta



Disertační práce

Výpočty parametrů spekter NMR a jejich užití ve strukturním modelování biomolekul

RNDr. Zuzana Vokáčová

Školitel: Dr. Vladimír Sychrovský

Konzultanti: Doc. RNDr. Ing. Jaroslav Burda, CSc., Prof. RNDr. Jiří Šponer, DrSc.

Praha 2009

To Jenda

„What do you get if you multiply six by nine?"

"Six by nine. Forty two."

"That's it. That's all there is."

"I always thought something was fundamentally wrong with the universe"

*(Answer to Life, the Universe, and Everything,
The Hitchhiker's Guide to the Galaxy)*

First of all I would like to thank my fiancé Jenda for his constant support, patience, and understanding without which this work would not be possible. Special thanks go also to my both families and my closest friends Luboš, Robert, and many other who helped me during my doctoral study.

I am grateful to my advisor Vladimír Sychrovský who guided me through a computational world of NMR parameters, to both my consultants Jaroslav Burda and Jiří Šponer for useful advice and comprehensive discussions about computational chemistry and RNA world.

Big thanks to Lucy, Lenka, Ivonne, Karla, Marge, Aneta, all of Sarkas, Electras, and Molly, Jacinta, Sona, Quinn, Eva, Ulrika, Lala, Ivonne, Jodi, June, Karla, Kaja, Halina, and Hanicka for dedicated not only computational time.

And last but not least, big thanks go to my colleagues and friends from Institute of Organic Chemistry and Biochemistry: Lád'a, Bohdan, Dan, Jakub, Honza, Jana Valery, Michal, Lucka, Petr M., Jarek, Zbyšek, Standa, Jirka Š., Jirka P., Jindra, Pavel, Helena, Mirka, Martin Š., Soňa, Martin D., Miloš B., and Petr B. for fruitful discussions and for very pleasant and enjoyable atmosphere.

I declare that I wrote this thesis myself and that it represents the results of my own work, unless stated otherwise in the text. All books, articles, internet sites, and other sources of information used are properly cited in the References section.

Neither the thesis nor any of its part has been used previously for obtaining any academic degree.

Zuzana Vokáčová

V Praze 30.6.2009

Contents

LIST OF FIGURES	7
LIST OF TABLE	7
LIST OF SCHEMES	7
LIST OF ABBREVIATIONS	8
 PREFACE	 9
 1. COMPUTATIONAL METHODS	 12
1.1 BASIC PRINCIPLES OF QUANTUM CHEMISTRY	12
1.2 CLASSICAL MOLECULAR MODELING	15
1.3 NUCLEAR MAGNETIC RESONANCE	18
 2 RESULTS	 22
2.1 DEPENDENCE OF NUCLEAR MAGNETIC PROPERTIES ON STRUCTURE OF NUCLEIC ACID BACKBONE	24
2.1.1 QUANTUM CHEMICAL CALCULATIONS OF J-COUPPLINGS	25
2.1.2 MOLECULAR DYNAMIC SIMULATIONS OF THE DMP MOLECULES	29
2.2 J-COUPPLINGS AROUND THE GLYCOSIDIC TORSION ANGLE IN GUANOSINE	33
2.3 TRANS-HYDROGEN J-COUPPLINGS IN WATSON-CRICK/SUGAR EDGE BASE PAIRING	35
 3 CONCLUSIONS	 40
 USED INSTRUMENTATION	 42
 REFERENCES	 43
 LIST OF PUBLICATIONS	 46
 ATTACHED PUBLICATIONS	 48

List of Figures

Figure 1: Chemical diagram of the RNA molecule.....	23
Figure 2: The RNA ApU dinucleoside monophosphate. The	24
Figure 3: Modulation of the calculated $^3J(\text{P-H5}')$ coupling [Hz] assigned to the torsion angle β [$^\circ$] due to variation of the neighboring torsion angle α	27
Figure 4: The calculated $^4J(\text{P-C3}')$ couplings [Hz] assigned to the torsion angle γ [$^\circ$]	28
Figure 5: Sketch of the interactions sides of the RNA mononucleoside.....	36
Figure 6: Geometry of cis-A.rA complex belonging to WC/SE base pair family.	36
Figure 7: The calculated mutual correlation of the $^3J(\text{C-N})$ and $^2J(\text{C-H})$ couplings [Hz] across the $\text{N-H}\cdots\text{O}=\text{C}$ link	38

List of Table

Table 1: The NA's backbone torsion angles and the J-couplings with which were correlated	26
Table 2: The observed conformations of the α - γ - ζ structural segment, their calculated population in MD time [%] and relevance of this triad to the stacking [%] calculated from time component.	31
Table 3: The calculated A, B, C parameters [Hz] and the phase ϕ [degree] of the Karplus equation for the $^3J(\text{C4-H1}')$ coupling.	34
Table 4: The calculated A, B, C parameters [Hz] and the phase ϕ [degree] of the Karplus equation for the $^3J(\text{C8-H1}')$ coupling.	34

List of Schemes

Scheme 1:	27
Scheme 2:	30

List of Abbreviations

A	adenosine
BLYP	Becke-Lee-Yang-Parr functional
B3LYP	Becke-Lee-Yang-Parr functional, 3 parameters
C	cytidine
CP DFT	coupled perturbed density functional theory
DFT	density functional theory
DMP	dinucleoside monophosphate
DNA	deoxyribonucleic acid
FF	force field
G	guanosine (dG – deoxy guanosine)
GGA	generalized gradient approximation
H-bond	hydrogen bond
HF	Hartree-Fock
LDA	local density approximation
NA	nucleic acid
NMR	nuclear magnetic resonance
MD	molecular dynamics
MM	molecular modeling
PBE	Perdue-Burke-Ernzerhof functional
PCM	polarized continuum model
QC	quantum chemistry
QM	quantum mechanic
RNA	ribonucleic acid
SE	sugar edge
U	uridine
WC	Watson-Crick

Preface

Nucleic acids (NAs) are molecules which play key role in many essential biochemical processes in living organisms. Although the number of basic building units in deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) molecules is limited only to four different nucleosides, the structural variation of their three-dimensional shapes found so far in nature is astonishing.¹ Structure and molecular flexibility of these biomolecules is usually closely related to their specific functions. Knowledge of the structural properties which can be nowadays acquired for the biomolecules with suitable experimental techniques and modern tools of the theoretical modeling is indispensable for detail understanding of many elementary biological processes; for example, the replication and transcription of the NAs or specific recognition of these molecules and their complexes by novel drugs in the living cells.

The famous DNA double helical structure was firstly resolved by James Watson and Francis Crick in 1953.² This essential discovery was soon followed by the structural studies of many different NAs which significantly contributed to the understanding of their structural variability.^{3,4} The detailed and accurate structural studies of these biopolymers are, however, still in the forefront of the contemporary science documenting thus large complexity of the task.

While the DNA structures are rather conservative since majority of the patterns observed in the nature include only three structural classes (A-, B- or Z-DNA), the RNA molecules possess in this respect truly extreme structural variations. This basic structural difference corresponds to the specific functionality of the DNA and RNA molecules. While storage and replication of the genetic information is domain of the DNA, translation and expression of the genetic information involves the RNA molecules. For a quite long time, the role of the RNA was underestimated and scientific research was massively focused mainly on the DNA. However, during the last three decades was the importance of the RNA molecules in regulation of gene expression recognized and current RNA research is rapidly expanding. Let us mention for example the research concerning RNA catalytic properties.⁵ Other recent studies unveiled many new and important aspects of RNA molecules supporting thus idea of the RNA critical role in regulation of the life processes usually called “the RNA world”.⁶⁻⁸

The annually increasing number of experimentally resolved NAs structures⁹⁻¹¹ reflects progress of the X-ray crystallography and nuclear magnetic resonance (NMR), the two major experimental techniques which are widely used for the molecular structure determination. Each of the methods has some advantages and disadvantages, but the final atomic model of a molecule established with both methods depends on the structural interpretation of measured experimental data. This thesis is focused on the interpretation of NMR spectroscopy parameters.

One of the significant advantages of the NMR spectroscopy is that it can be used for probing the molecular structure in solution what, contrary to the X-ray, may correspond better to the native environment of biomolecules.¹²⁻¹⁴ The molecular structure determined with the NMR may utilize different spectroscopy parameters, for example: the nuclear Overhauser effect (NOE), the chemical shift including its anisotropy, the indirect spin-spin coupling constant (J-coupling) and the cross-correlation relaxation rates.^{12,15-18} The NMR parameters generally possess different dependence on local molecular geometry, although all can provide valuable information about the molecular conformation. For example, the chemical shifts of ^1H , ^{13}C , ^{15}N , and ^{31}P nuclei are particularly sensitive to the chemical bonding of a probed atom. The structural information deduced from the chemical shifts is therefore pretty much local. On the other hand, the J-couplings can be considered in this respect as a source of the long-distance structural information since they are determined along a chain of atoms which can be bound either covalently¹² or non-covalently, i.e. across the hydrogen bond^{13,19-21}. Especially the lastly mentioned type of the intermolecular interactions is fundamental for the folding properties of NAs and their detection with the NMR can significantly improve the quality of structural model.

The NMR spectroscopy is still limited with respect to the size of probed molecules despite recent technical advances. The preparation of a molecular sample therefore may include the site specific labeling with relevant nuclei in order to overcome this problem, but the cost of such labeled sample can increase dramatically. The structural interpretation of measured NMR parameters may become problematic when the effects of local dynamics or specific interactions of molecular sample with solvent complicate the structural assignment. Another complication could be the multidimensional dependence of the NMR parameters, i.e. when the dependence of the NMR parameter on the assigned geometry coordinate is not stable with respect to a variation of some other coordinate(s) of the conformational space.^{22,23} It is therefore not surprising that the dependence of the NMR parameters on the molecular structure of the NAs has been so far parameterized mostly empirically (i.e. on the basis of

measured data for a priory known molecular structure) and only for rather limited number of the spectroscopic parameters.¹² The calculations with methods of computational chemistry can therefore reliably complement the rules for the structural interpretation of the NMR parameters which are measurable in the NAs.^{24,25} Further, the theoretical modeling of the NMR parameters carried out for properly selected part of measured macromolecule was shown to be well suited methodology since the obtained rules can be often applied generally for larger systems due to the local character of the magnetic interactions.

The methods of computational chemistry used in this doctoral work, i.e. the quantum mechanics (QM) and the molecular dynamics (MD), have nowadays become standard techniques which often complement detail analysis of the experimental data.

This doctoral thesis is focused on the theoretical calculation of the NMR parameters, in particular on the structural interpretation of NMR spectra measured in NAs and also on the prediction of new parameterizations and structural dependences for some selected NMR parameters with the goal to improve the currently used interpretation schemes.

Chapter 1 contains briefly survey of the used methodology concerning the quantum chemistry methods and principals of molecular dynamic simulations including the basics of theoretical NMR spectroscopy focused on selected NMR parameters.

Chapter 2 and 3 include presentation of the essential results, and conclusions obtained during my doctoral studies.

The articles which were published during the doctoral study are explicitly attached in the section Attached Publications.

1. Computational Methods

1.1 Basic Principles of Quantum Chemistry

The Non-Relativistic Time-Independent Schrödinger's equation

The main aim of quantum chemistry²⁶⁻²⁸ is solution of the non-relativistic time-independent Schrödinger's equation

$$\hat{H}\Psi = E\Psi, \quad (1.1)$$

which means determination of both the wave function Ψ describing state of a molecular system and its total energy E . The operator \hat{H} is the Hamilton operator which includes in the non-relativistic limit sum of the five terms

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn} \quad (1.2)$$

corresponding to the kinetic energy operator of nuclei (\hat{T}_n) and electrons (\hat{T}_e), the potential energy between electrons and nuclei (\hat{V}_{en}) and the potential energy of electron-electron (\hat{V}_{ee}) and nuclei-nuclei (\hat{V}_{nn}) repulsion. These terms can be written as

$$\begin{aligned} \hat{T}_n &= -\sum_{A=1}^M \frac{\hbar^2}{2M_A} \nabla_A^2, & \hat{T}_e &= -\sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2, & \hat{V}_{ee} &= \sum_{i=1}^N \sum_{j>i}^N \frac{e^2}{r_{ij}} \frac{1}{4\pi\epsilon_0}, \\ \hat{V}_{en} &= -\sum_{i=1}^N \sum_{A=1}^M \frac{Z_A e^2}{r_{iA}} \frac{1}{4\pi\epsilon_0}, & \hat{V}_{nn} &= \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B e^2}{R_{AB}} \frac{1}{4\pi\epsilon_0}, \end{aligned} \quad (1.3)$$

where i, j and A, B are indices of N electrons and M nuclei, respectively. M_X and Z_X are the mass and atomic number of nucleus X , r_{xy} is the distance between electrons x and y , R_{XY} is the distance between nuclei X and Y and r_{xX} is the distance between electrons x and nucleus X .

Born-Oppenheimer Approximation

The Born-Oppenheimer approximation is an almost universally assumed approximation in quantum chemistry which basically utilizes the fact of different mass of electron and nuclei. The total wave function depending on the coordinates of both electrons and nuclei can be under such prerequisite approximated as product of an electronic and nuclear component, i.e. the motion of electrons and nuclei were in fact separated.

$$\Psi_{total} = \Phi_{elec} \cdot \Phi_{nucl} \quad (1.4)$$

The Born-Oppenheimer approximation can be also deduced from the respective kinetic energy operators (Equations 1.3) by considering the mentioned mass effect. The electronic Schrödinger equation yielding the Φ_{elec} wave function can be thus within the Born-Oppenheimer approximation solved separately and the nuclear coordinates are treated as parameters. The nuclear coordinates corresponding to the minimum of total energy E called the optimal geometry can be found using some numerical algorithm on the potential energy surface which is calculated as dependence of the total energy E on nuclear coordinates.

Hartree-Fock Method

The Hartree-Fock (HF) method is the most widespread ab-initio method for the solving the Schrödinger equation (1.1). It assumes the representation of the electronic wave function Ψ for the N -particle fermionic system in a form of single Slater determinant fulfilling thus the anti-symmetry requirements put on the electronic wave function. The HF equations actually form the set of N equations, each depending only on coordinates of a single electron. The HF method can be therefore characterized as a one-electron approximation for solving the general Schrödinger equation since each electron is subjected to the mean field created by all other electrons. The HF equations are solved iteratively using self consistent field approach. Computationally convenient approximation of true electron-electron interactions by the mean field introduced for one electron in the HF equations, however, results in incomplete description of the electron correlation effects. This deficiency can be more or less cured with the so called post-HF computational methods which include the electron correlation, but their applicability can be in some cases limited only to relatively small systems.

Density Functional Theory

The density functional theory (DFT) represents theoretical approach which is similar to the HF method, but the DFT method includes the electron correlation. In the DFT the N -electron system is described by an electronic density $\rho(r)$, instead of the wave function Ψ . Hohenberg and Kohn proofed that the ground state energy is uniquely determined by the electron density $\rho(r)$.²⁹ Replacement of the wave function by the electronic density is computationally very convenient, but the mathematical form of the electron density which is called the exact DFT functional is not yet known. However, the functional of total energy, i.e. the mathematical form which for given density $\rho(r)$ gives the total energy E , can be always written in this form

$$E[\rho] = \int \rho(r) v(r) dr + T[\rho] + V_{ee}[\rho], \quad (1.5)$$

where the first term describes the interaction of the electronic density with an external potential $v(r)$, the second term is the kinetic energy and the third term is the functional describing the electron-electron Coulomb repulsion.

Major task of the DFT developers is to find accurate approximation to the exact functional. This problem was addressed by Kohn and Sham who actually suggested introducing HF-like wave functions to equation (1.5), used Hohenberg-Kohn theorem²⁹ and rewrote it as a set of N one-electron equations in full correspondence with the HF method. Kohn-Sham's equations then contains only one term which was unknown – the exchange-correlation potential V_{XC} , which cover exchange and correlation effects and which is not included in other terms. Exact V_{XC} is not known, therefore approximation are used.

Because the number of currently available DFT functionals is nowadays truly large we would like to mention here only the three following classes. The local DFT functionals depend only on the electron density (LDA – local density approximation) while the non-local DFT functionals depend also on derivatives of the density (GGA - generalized gradient approximation, e.g. BLYP or PBE functional). Actual overall performance of the pure DFT functional suggested so far was usually not satisfactory. Further improvement of the two mentioned DFT functionals was achieved by mixing the pure DFT approach with the HF theory which led to the so called hybrid functionals (e.g. B3LYP).

In this doctoral thesis the DFT approach was extensively used namely for obtaining the optimal molecular geometries and NMR parameters which were calculated mostly with the B3LYP functional. The geometry optimization was mostly done with the 6-31G** basis set while the calculations of the NMR parameters were done with the atomic basis Iglo II and Iglo III which were specially designed for NMR properties calculation. The DFT calculations were done with the Gaussian 03 program package.³⁰

1.2 Classical Molecular Modeling

The methods of classical molecular modeling (MM) are based on time propagation of particles in space on the basis of the Newton's laws. The MM methods represent a powerful tool for simulating the structural properties of large molecular systems with full biochemical relevance that can be otherwise hardly calculated with the first principle QM methods. Total energy of a simulated system in each step of the MM calculation depends critically on the force field (FF) parameters. The FF generally models different kinds of forces acting on particles involved in specific type of bonding in a molecule. The FF parameters therefore particularly correspond to the stretching of bonds including variation of bond and torsion angles as will be discussed below. This approximation concerning the basic structural behavior of molecules considerably reduces the computational requirements for calculating the dynamical properties of many-particle systems which can be even embedded in the explicit solvent. On the other hand the quality of the FF is probably the most critical point as regards the accuracy and reliability of the simulated behavior of molecules.

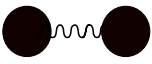
Molecular Mechanics

Molecular mechanics is the method for calculating energy of a system for given configuration of particles interacting via FF. It can be used for obtaining the optimal molecular geometry and for calculating energy and its derivatives in every time step of MM simulation. The total energy V of a system is calculated as a sum of contributions describing the potential energy of bond stretching, angle bending, bond rotation, and other terms corresponding to the Van der Waals and Coulombic interactions.

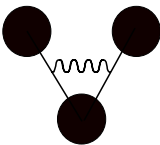
$$V = \underbrace{\sum_{\text{Bonds}} E^{\text{stretch}} + \sum_{\text{Angles}} E^{\text{bend}} + \sum_{\text{Torsions}} E^{\text{torsion}}}_{\text{BONDED}} + \underbrace{\sum_{i < j} E_{i,j}^{\text{VDW}} + \sum_{i < j} E_{i,j}^{\text{EL}}}_{\text{NONBONDED}} + \dots \quad (2.1)$$

Individual terms of the potential energy V corresponding to the bonding and non-bonding interactions can be approximated by the following simple analytical functions.

Analytical functions describing the bonding terms of the potential energy V

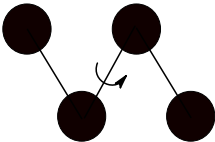


$$E_i^{\text{stretch}} = \frac{k_i}{2} (r_i - r_{i,0})^2$$



$$E_i^{\text{bend}} = \frac{l_i}{2} (\Theta_i - \Theta_{i,0})^2$$

(2.2)



$$E_i^{\text{torsion}} = \frac{V_i}{2} \left(1 + \cos \left[n (\omega_i - \omega_{i,0}) \right] \right)$$

Magnitude of the force constants k_i and l_i reflects stiffness of bond and angle bending, respectively. V_i is the barrier of the torsion rotational potential with periodicity n . $r_{i,0}$, $\Theta_{i,0}$ and $\omega_{i,0}$ are the equilibrium values of bond length, angle and torsion angle, respectively. All these FF parameters are either fitted to some experimental data or calculated with some QM method. The r_i , Θ_i and ω_i variables of the potential energy corresponding to the bond length, valence angle and dihedral angle. These basis FF parameters are often complemented by other parameters describing for example the deviations of one atom from the plane defined by other three atoms called the improper torsion angle.

Analytical functions describing the non-bonding terms of the potential energy V

$$E_{i,j}^{\text{EL}} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.2)$$

$$E_{i,j}^{\text{VDW}} = 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

The Coulomb term describes the interaction between two particles with charges q_i and q_j where ϵ_0 is the permittivity of vacuum. The Van der Waals interaction is usually describe by Lennard-Jones potential where ϵ is the depth of the potential, σ_{ij} is the distance between atoms i and j (at which the interatomic potential is zero), r_{ij} is the distance between atoms i and j .

Molecular Dynamics

The method of molecular dynamics (MD) is used for calculation of dynamical behavior of molecules in time including calculation of the thermodynamic properties like pressure, temperature and volume of particles in modeled system. The evolution of particles in time is described with the MD equations which are based on the classical Newton's equations of motion. A force \vec{F}_i acting on a particle i can be computed directly from the derivative of the potential energy V with respect to the coordinates r_i

$$\frac{d^2 \vec{r}_i}{dt^2} = \frac{\vec{F}_i(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)}{m_i} = -\frac{1}{m_i} \frac{\partial V}{\partial \vec{r}_i}, \quad (2.3)$$

where m_i is mass of a particle and r_i is its position. The equations (2.3) are solved with some numerically algorithm. All numerical algorithms are based on Taylor expansion of coordinates, velocities and forces in time. The initial velocities and accelerations in the first step of MD simulations are set randomly according to the Maxwell-Boltzmann statistics for given temperature. The coordinates, velocities and accelerations in each step are calculated from their magnitudes in the previous step. The set of coordinates and velocities obtained during computation is called a trajectory.

The statistical ensembles used for the MD description of many-particle system may differ depending on which of the state variables is kept fixed during computation. Among the most commonly used are the NPT (constant number of atoms/ions, pressure and temperature), NVT (constant number of atoms/ions, volume and temperature) and NVE (constant number of atoms/ions, volume and energy) ensemble. Particular choice of the statistical ensemble depends on character of the problem. For example, in projects including in this doctoral thesis we used the NPT statistical ensemble in MD simulations for the dynamical properties of RNA dinucleoside monophosphates (DMP) at room temperature. The structure of the RNA molecules was relaxed and equilibrated in the solvent modeled with explicit water molecules

prior the 50 ns MD simulations carried out with the Amber Molecular Dynamics Package, version 8,³¹ using the AMBER parm99³² FF.

1.3 Nuclear Magnetic Resonance

The nuclear magnetic resonance spectroscopy (NMR) is the experimental spectroscopy technique which probes the relaxation properties of atomic nuclei having the non-zero nuclear spin by applying the external magnetic field. Since the local magnetic field experienced by the nuclear magnetic moments reflects specifically the chemical bonding and structural arrangement of atoms in molecules the NMR spectroscopy can be used for determining the molecular structure. NMR spectroscopy is probably the only spectroscopy which currently allows real atomistic resolution of molecules in liquid phase what is indispensable for the biomolecular research.

The Nuclear Spin

Relaxation properties of the nuclear spin are in a forefront of experimental NMR spectroscopy. Various well elaborated NMR pulse sequences must be applied in order to measure different NMR parameters. The physical model of NMR measurement basically corresponds to the probing of nuclear spin quantum states with the external magnetic field. Since the scope of this doctoral thesis is limited to the calculations of NMR parameters we would like to mention only some of the important aspects of experimental NMR which are anyway related to the topic because our calculations were done also in joint collaboration with experiment.

The nuclear spin of different atoms and their isotopes can be either integer or half integer positive number. The nuclei with spin $\frac{1}{2}$ are of major importance in the experimental NMR spectroscopy because their spectral pattern is the simplest pattern which can be achieved. The NMR assignment of such spectra is relatively easy since other nuclei with the larger spins can make interpretation of NMR spectra less straightforward due to the complicated multiplet patterns which usually overlap. Biomolecules consist mostly of the hydrogen, carbon, nitrogen, oxygen and phosphorus atoms. Natural abundance of the ^1H and ^{31}P isotopes is 100%, but in the case of ^{13}C and ^{15}N isotopes it is only 1.1% and 0.4%, respectively. The difficulties with sensitivity of molecular sample can be overcome by its labeling with relevant isotopes, but actual price of the labeled molecules increases.

Let us finally mention that the QM computations of NMR parameters are not based on any theoretical model manipulating with the nuclear spin as actually happens in the NMR experiment by irradiating molecular sample with the external magnetic field. The magnetic field produced by electrons which can be modeled in a framework of QM methodology and the nuclear magnetic moment or the external homogenous magnetic field are treated as perturbation parameters are however in full correspondence with the physical basic of the NMR.

Chemical Shift

Total magnetic field at the site of the nuclear spin including the external homogenous magnetic field, the magnetic field created by circulating electrons and the field produced by other neighboring nuclear spins basically affect the relaxation properties of the nuclear spin. Different kind of the chemical bonding or geometry changes resulting in different local magnetic field created by variation of the electronic structure makes the shift in frequency of the nuclear spin resonance. We say in such case that the nuclear spin is shielded differently in the different local magnetic fields. The chemical shift is calculated as difference between the shielding of the reference and actually probed nucleus.

The chemical shielding $\vec{\sigma}$ is the tensor property since it couples the two vectors of the homogenous magnetic field \vec{B}_0 and the nuclear magnetic moment \vec{I} to the energy contribution. The components of the chemical shielding tensor can be calculated with the QM methods as the second partial derivative of the total energy including the magnetic interactions described by the magnetic Hamiltonian H_m with respect to the \vec{B}_0 and \vec{I} .

$$\vec{\sigma} = - \frac{\partial^2 \langle \Psi | \hat{H}_m | \Psi \rangle}{\partial \vec{I} \partial \vec{B}_0} \quad (3.1)$$

The trace of shielding tensor is invariant to any translation or rotation of the coordinate system. The isotropic chemical shielding is defined as a one third of this invariant

$$\sigma_{iso} = \frac{1}{3} (\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) \quad (3.2)$$

where σ_{ii} are the diagonal elements of the shielding tensor, corresponds to the value averaged over the isotropic motion of a molecule. The calculated isotropic value of chemical shift can be directly compared with the value measured in liquid phase, however, the individual components of the shielding tensor can be also measured in the solid state NMR experiment.

Indirect Spin-spin Interaction

Interaction of two nuclear spins mediated by the electronic system of the molecule is called the indirect spin-spin coupling or the J-coupling. In NMR experiment such interaction causes the line splitting into doublets, triplets or even higher multiplets in dependence of multiplicity of the two-spin system.

The J-coupling is also tensor property since it couples the \vec{I}_j and \vec{I}_k spins and can be calculated as the second partial derivative of the total energy including the magnetic interactions.

$$\vec{J}_{jk} = - \frac{\partial^2 \langle \Psi | \hat{H}_m | \Psi \rangle}{\partial \vec{I}_j \partial \vec{I}_k} \quad (3.3)$$

Similarly as the isotropic NMR shifts the isotropic value of the J-couplings (calculated in the same way) can be directly compared with the J-couplings measured in liquid phase. Magnitude of the J-couplings can provide useful information about the molecular structure along the spin-spin coupling pathway. This is often utilized in studies dealing with the molecular conformations. The three-bond J-coupling (3J) assigned to the respective torsion is frequently used determining the angle magnitude. The dependence of 3J -couplings on torsion angle was firstly analyzed by Martin Karplus in 1963 on the basis of series of the Fourier coefficients truncated after the third term.³³

$$^3J(X-Y, \theta) = A \cos^2(\theta - \varphi) + B \cos(\theta - \varphi) + C \quad (3.4)$$

Such parameterization of the 3J -couplings between atoms X and Y depending on torsion angle θ , where A , B and C are the amplitudes and φ is the phase shift is called the Karplus equation.

For the J-couplings between atoms separated by n bonds is used the $^nJ(X-Y)$ notation, while for the spin-spin interaction of atoms linked by a non-covalent contacts is usually used the $^{nh}J(X-Y)$ notation.

The NMR spectroscopy parameters in this doctoral work were calculated using the coupled perturbed DFT (CP-DFT) method^{25,34,35} with the B3LYP DFT functional. This method was found previously to possess the best overall performance among the other commonly used DFT functionals. The two atomic basis sets Iglo II and Iglo III³⁶ designed specially for calculating the NMR properties were used. Water solvent effects on NMR parameters were modeled with the polarized continuum model (PCM).³⁷ The calculations were carried out with the mentioned methods as they are implemented in the Gaussian 03 program package.³⁰

2 Results

Main topic of this doctoral thesis concerns the calculation of NMR parameters in RNA molecules carried out for the individual NA bases, dinucleoside fragments of RNA molecules and also for the nucleobases interacting via hydrogen bonds with nucleosides. Other subprojects treated during the doctoral studies are shortly listed at the end of this paragraph.

The RNA molecules are similar to the DNA ones as regards the composition of basic unit – the nucleoside. Both molecules actually contain the same backbone part linking the sugar units of the two neighboring nucleosides which is shown in Figure 1. The two structural differences between the RNA and DNA molecules are: 1) the 2'-hydroxyl group of the ribose sugar in RNA while DNA has the deoxy-ribose sugar, 2) different set of nucleobases in RNA (adenine, guanine, cytosine and uracil) and DNA (adenine, guanine, cytosine, thymine). Schematic sketch of all RNA bases is shown in Figure 1 including the numbering of atoms and definition of the torsion angles. The mentioned differences in composition of RNA and DNA molecules induce large conformational variability of RNA molecules in comparison with DNA molecules.

The results of this doctoral thesis are presented in separate chapters as follows.

The first part contains the results of QM calculations of J-couplings in the sugar-phosphate backbone of RNA and DNA dinucleosides carried out for a wide variety of the NA backbone conformations which were found in crystals of ribosomal RNA.²³ Five different ribosomal structural classes of backbone were further used as the starting patterns for MD calculations of the ApA, ApC, CpA and CpC RNA dinucleosides in order to model their specific dynamic behavior.³⁸ The dynamical structural information obtained for the backbone and glycosidic torsion angles were used for calculating of the dynamical averages magnitude of J-coupling including the molecular flexibility.

In the second part we present the results of QM calculations focused on correlation of the ^3J and ^1J -couplings with geometry parameters of the glycosidic torsion angle in guanosine.³⁹ New parameterizations of Karplus equations included several structural and solvent effects.

The third part of the Results presents trans-hydrogen bond J-couplings calculated in the complexes of RNA nucleobases and nucleosides belonging to the *cis* and *trans*-Watson-Crick/Sugar Edge base pair family.²¹

Other subprojects of the doctoral studies were also focused on calculations of the molecular properties; in particular, the calculations of the ¹³C and ¹⁵N magnetic shielding anisotropy in DNA nucleosides⁴⁰ spectral properties of the selected pigments in various photosystems⁴¹ and dependence of NMR parameters in the L-alanyl-L-alanine peptide²². The corresponding publications are also attached to this thesis.

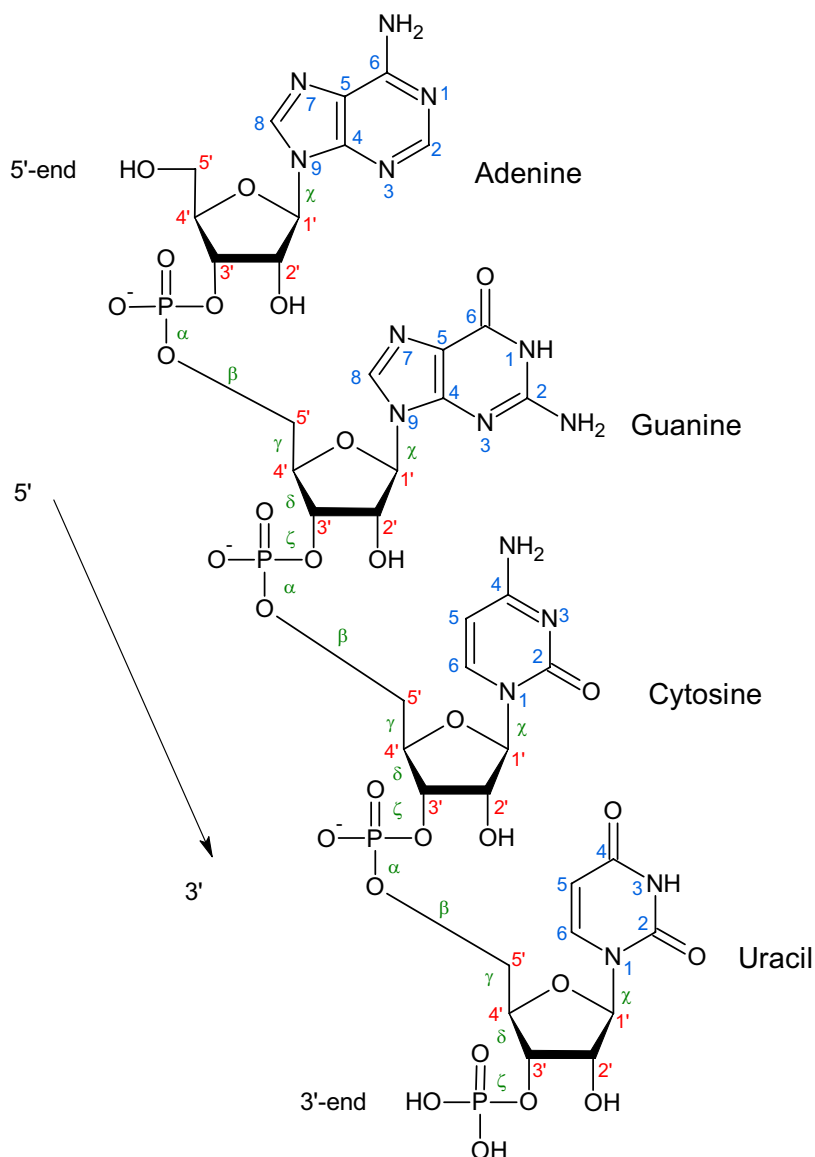


Figure 1: Chemical diagram of the RNA molecule with the indicated 5'→3' direction of the backbone and numbering of atoms in the sugar-phosphate backbone (red) and nitrogenous bases (blue). The torsion angles are defined in the standard way (green Greek letters): $\alpha = O3'-P-O5'-C5'$, $\beta = P-O5'-C5'-C4'$, $\gamma = O5'-C5'-C4'-C3'$, $\delta = C5'-C4'-C3'-O3'-P$, $\epsilon = C4'-C3'-O3'-P$, $\zeta = C3'-O3'-P-O5'$, $\chi = O4'-C1'-N1-C2$ for the pyrimidine base (cytosine and uracil) and $\chi = O4'-C1'-N9-C4$ for the purine base (adenine and guanine).

2.1 Dependence of Nuclear Magnetic Properties on Structure of Nucleic Acid Backbone

The 3'-5'-dinucleoside monophosphate (DMP, N-p-N, where N is nitrogenous base, p is phosphate link, Figure 2) was chosen as a reliable model compound for this theoretical study of conformational dependence of NMR parameters in the NA backbone. From the structural point of view, the DMP represents the smallest fragment of RNA that has still the key chemical and structural features of the NAs. The repetitive DFT calculations of NMR parameters for this quite large molecular fragment can be done still in reasonable time. Structure parameters of the DMP molecule (Figure 2) include all torsion angles used for description of NA backbone (Figure 1).

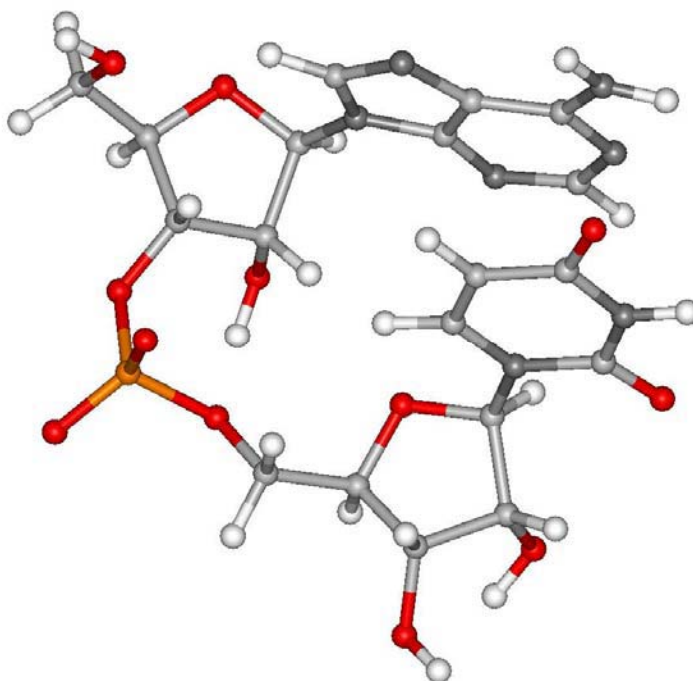


Figure 2: The RNA ApU dinucleoside monophosphate. The 5'-end and 3'-ends of the DMP were terminated by hydroxyl group at carbon C5' and carbon 3', respectively (numbering of atoms are shown in Figure 1).

2.1.1 Quantum Chemical Calculations of J-couplings

The J-couplings were calculated between the ^{31}P , ^{13}C and ^1H nuclei with DFT method (B3LYP, IgloIII, PCM implicit solvent model, for the details see Ref^{23 16}) in 16 representative conformers of the RNA backbone and the J-couplings were assigned to respective backbone torsion angles. Interestingly, the conformations of backbone in ribosomal RNA has been shown to cluster only in relatively small number of classes,^{42,43} so the systematic scanning of the conformational space of backbone torsion angles was not necessary.. The backbone of the DMP is described by seven torsion angles (Figure 1 and 2): α , β , γ , ϵ , ζ and δ torsion angle of the two ribose units. The selected RNA conformational classes (Table 1 in Ref²³) cover all important backbone topologies with high populations found in the tracked ribosomal subunit.⁴² The selected RNA classes used for the calculation of NMR parameters cover also different mutual topologies of nucleobases including the bases in stack geometry, parallel bases, perpendicular bases, bases lying in platform, and the completely open structure without any overlap of nucleobases. The twenty different J-couplings calculated cover majority of the J-couplings which can be measured in the NA backbone. Assignment of the modeled J-couplings includes all backbone torsion angles (Table 1). This theoretical study showed that the backbone J-couplings can be accurately and reliably interpreted since the theoretical data basically agree with number of measured J-couplings available in the literature. Further, the set of relevant J-couplings determined along the NA backbone can be used for discriminating among different conformational classes, i.e. each individual RNA backbone conformation posses its specific “NMR fingerprint”.

Geometry of the structural models taken from X-ray database was firstly relaxed using the optimization procedure with the Amber program package with fixed backbone torsion angles in order to keep the overall conformation of the NAs classes. The J-couplings were calculated using the CP-DFT method with the B3LYP functional and with atomic basis set Iglo II and Iglo III.

Table 1: The NA's backbone torsion angles and the J-couplings with which were correlated (for atom numbers and labeling of torsion angles see Figure 1). Numerical index of the J-couplings indicates a number of the bonds between the coupled nuclei.

NAs torsion angle	J-coupling
α , O3'-P-O5'-C5'	$^2J(P-C5')$
β , P-O5'-C5'-C4'	$^3J(P-C4')$, $^3J(P-H5')$, $^3J(P-H5'')$
γ , O5'-C5'-C4'-C3'	$^3J(H4'-H5')$, $^3J(H4'-H5'')$, $^3J(C5'-H3')$, $^2J(C5'-H4')$, $^2J(C4'-H5')$, $^2J(C4'-H5'')$, $^2J(C5'-H3')$, $^4J(P-H4')$, $^4J(P-C3')$
δ , C5'-C4'-C3'-O3'	$^3J(H3'-H4')$, $^3J(H2'-H3')$,
ϵ , C4'-C3'-O3'-P	$^3J(P-C4')$, $^3J(P-C2')$, $^3J(P-H3')$
ζ , C3'-O3'-P-O5'	$^2J(P-C3')$
χ , O4'-C1'-N1/9-C2/4	$^3J(C6/8-H1')$, $^3J(C4/2-H1')$

At first, according to our results, the classification of the torsion angles were done. The distribution functions calculated with MD for backbone torsion angles correspond well to the distributions resolved with X-ray in ribosomal RNA.⁴² Hence, the backbone torsion angles in the DMP can be divided according to the shape of their distributions into two groups: a) torsion angles with bi- and trimodal narrow distributions ("sharp" torsion angles; α , γ , δ), and b) torsion angles with wide distributions ("soft" torsion angles; β , ϵ , ζ). This classification of the torsion angles with notation "soft" and "sharp" also used in our publication²³ reflects the width of rotational potential and stiffness of torsion angle rotation.

The different shape of the distributions calculated for backbone torsion angles showed to have an effect on magnitude of the calculated J-coupling. It was mostly exhibit in splitting of the calculated magnitudes of the J-couplings (Figure 3). For one magnitude of torsion angle J-coupling differs about 1 Hz in some regions. Dependence of the J-couplings on the structure can be then describing by more Karplus curves. This multidimensional dependence on the given torsion angle which was calculated for several J-couplings: γ , ϵ and β (see Figure 2 in Ref.²³). For example, the calculated J-couplings differ by 1-1.5 Hz for the $^4J(P-C3')$ and $^3J(C5'-H3')$ couplings and 2.3 Hz for the $^3J(P-H5')$ coupling (Figure 2 in Ref.²³, Figure 3), assigned to the torsions γ and β . Their dependence on the assigned torsion angle is systematically perturbed by the variation of neighboring torsion angles – the torsion angle β was affected by the torsion angle α , and the torsion angle γ was affected by the torsion angle δ which had strong correlation with sugar pucker conformation.

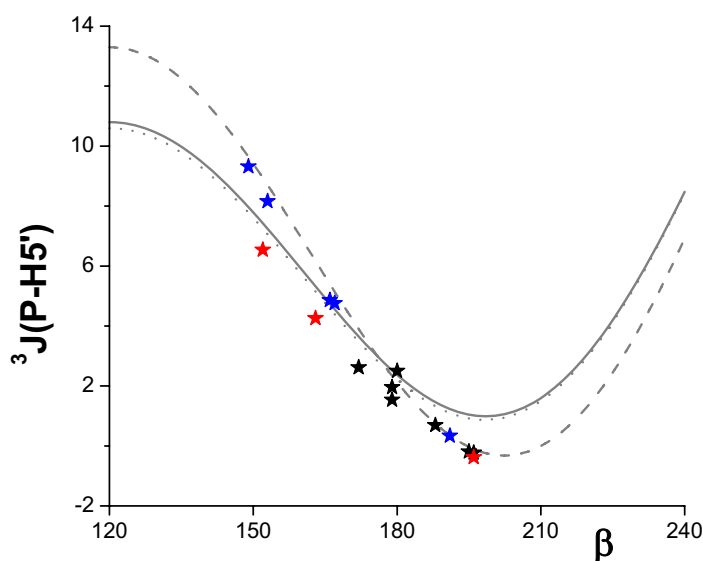
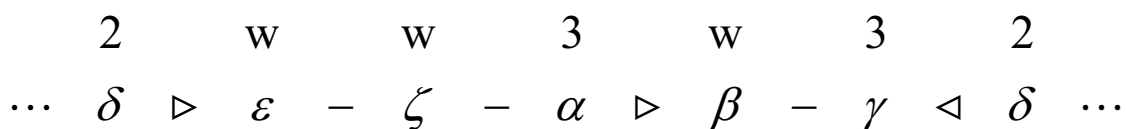


Figure 3: Modulation of the calculated ${}^3J(\text{P-H5}')$ coupling [Hz] assigned to the torsion angle β [°] due to variation of the neighboring torsion angle α . Calculated for the three amplitudes of α torsion angle: black ($\alpha \approx 70^\circ$), red ($\alpha \approx 160^\circ$), and blue ($\alpha \approx 300^\circ$) stars including the empirical Karplus curves: solid (Ref.¹⁵), dashed (Ref.⁴⁴), and dotted (Ref.¹²) line.

Generally, each of the “sharp” torsion angles embodies non-negligible affecting to the neighboring “softer” torsion angle. The variation of the “sharp” torsion angle affected the magnitudes of the J-couplings and modulates the Karplus curve assigned to the “soft” torsion angle (Figure 2 in Ref.²³, Figure 3). According to this mutual torsional affection, a protocol for a step-by-step assignment of the torsion angles in the NA backbone from the J-coupling measurements was proposed (schematically in Scheme 1). The “soft” torsion angles are labeled by letter w as they have wide torsional distribution, and the “sharp” torsion angles are termed by a number which describes how many peaks occurring in their structural distributions (2: bimodal, 3: trimodal). Arrows in the scheme indicate the direction of the neighboring affecting. According to the aforementioned protocol the “sharper” torsion angles should be assigned first.

Scheme 1:



In addition, this influence of the “sharp” torsion angle is stepped. Although torsion angle γ is also from the “sharp” group, it is modulated by the torsion angle δ , whose distribution is narrower than γ distribution. The effect of the torsion angle β is shown here for the $^4J(\text{P-C3}')$ coupling (Figure 4).

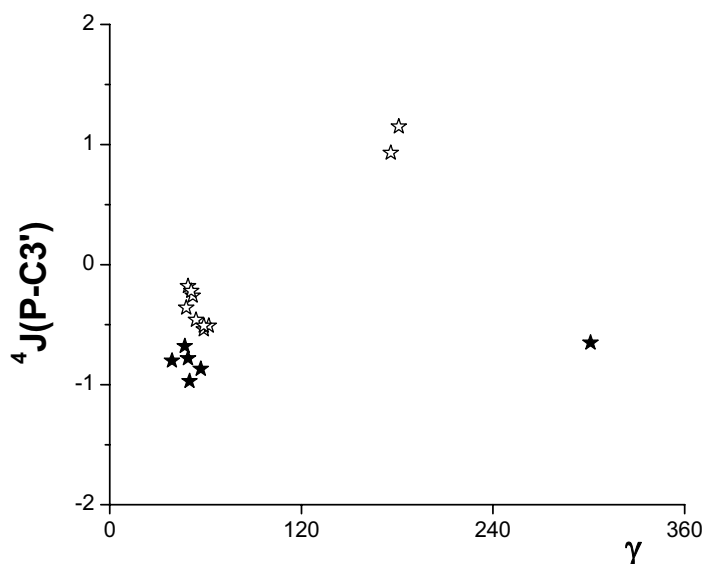


Figure 4: The calculated $^4J(\text{P-C3}')$ couplings [Hz] assigned to the torsion angle γ [°]. The clustering of calculated J-couplings correlates with two different magnitudes of the neighboring torsion angle δ which corresponds to the sugar pucker, empty stars ($\text{C3}'\text{-endo}$) and filled stars ($\text{C2}'\text{-endo}$).

For each torsion angle, at least one J-coupling was calculated however several J-couplings were mostly calculated for one torsion angle. For some of them, our calculations showed other possibility how to determine the torsion angle region. For example, the three main intervals of γ torsion can be determined only on from the sign of relevant J-couplings: see the Table 6 including the signs of the $^2J(\text{C4}'\text{-H5}')$, $^2J(\text{C4}'\text{-H5}'')$, $^2J(\text{C5}'\text{-H4}')$, $^4J(\text{P-C3}')$ and $^4J(\text{P-H4}')$ couplings in Ref²³.

The theoretical data obtained for the 2J -couplings assigned here to the torsion angles α and ζ showed the possibility for the determination of their magnitudes since the α and ζ torsion angles cannot be otherwise determined due to the lack of nuclei relevant for NMR measurement (Figure 1). The calculated $^2J(\text{P-C5}')$ and $^2J(\text{P-C3}')$ couplings were correlated with the α and ζ torsion angles. All possible combinations of the three conformational regions

available for each torsion angle were explored. The calculation showed that the $^2J(\text{P-C5}')$ and $^2J(\text{P-C3}')$ couplings can be used for discriminating between the gauche ($\pm 60^\circ$) and trans (180°) conformers due to the change of the 2J -couplings ranging 2-3 Hz. Together with the commonly accepted constrain excluding simultaneous occurrence of the α and ζ torsion angle in the trans region the calculated data showed the possibility for better detection of the backbone conformations in the vicinity of the phosphate group.

The calculated J-couplings correlate well with available experimental data from the literature.⁴⁵⁻⁵⁹

The effect of solvent on calculated NMR parameters was calculated with several models of aqueous solvents in the B-DNA DMP molecule. This backbone pattern possess well characterized first hydration shell in the solid state which was used for modeling the explicit hydration shell of the phosphate group which consisted of six water molecules. The performance of the explicit hydration was fully comparable with the PCM one (the calculated differences of all J-couplings were smaller than 2.2 Hz). This fact validates good overall performance of the PCM hydration model which was actually used calculations of the J-coupling in RNA backbone patterns.

Although the NMR properties are generally considered to possess local nature we performed the test calculations in order to estimate the effect of nitrogenous bases on the calculated backbone J-couplings. The calculated backbone J-couplings without and including nitrogenous bases differed less than 0.5 Hz what validates replacement of nitrogenous bases by methyl groups in actual calculations.

The effect of size of the atomic basis set on the J-couplings was estimated for the method including PCM hydration and the structural model with nitrogenous bases substituted by methyl groups. The difference between the J-couplings calculated with Iglo II and Iglo III basis were smaller than 0.6 Hz, so the effect of atomic basis is comparable with the error due to the substitution of the nitrogenous bases by the methyl groups and both effects on J-coupling magnitudes are smaller than the hydration effect.

2.1.2 Molecular Dynamic Simulations of the DMP Molecules

To understand the function of the RNA molecules and determine the RNA folding, the dynamical information is needed. By including the molecular dynamic in description of the NA backbone the interpretation of measured J-couplings also brings this important aspect of

molecular flexibility. The dynamical information concerning the RNA backbone was captured using the classical MD calculation method.

The MD calculations were performed for ApA, ApC, CpA, and CpC RNA molecules. We note that the 3'- and 5'-end of the DMP molecules are not equivalent (Figures 1 and 2). The four RNA dinucleosides are therefore structural models which are relevant for the study of relations between various backbone topologies and base stacking due to the complete sequential permutations of the purine (adenine (A)) and pyrimidine (cytosine (C)) nucleobases. The combinations of the A and C nucleobases were selected on purpose because their different ability for the base stacking can be expected.

For each DMP molecule we selected five different structures as starting structures in subsequent MD calculations in order to sample better the conformation space, two A-RNA classes and three unstacked DMP conformers. The starting structures in MD simulations cover variation of all important structural features including the backbone angles, base-stacking arrangement and sugar pucker conformation. All twenty initial models (5 different conformers combined with four different sequences) were hydrated in the box of explicit water molecules modeled with TIP3P model of water, equilibrated using the standard procedures with Amber program package. The MD calculations were carried out with the periodic boundary conditions, temperature of 300K and cut off 10 Å (For more details see Ref³⁸). Simulation time was 50ns for each of the five patterns of the four DMP molecules.

The backbone torsion angles in MD simulations possess exchange among their typical amplitudes occurring on time scales which are specific for each torsion and motions of some torsion angles are coupled. The typical time scales can be related to character of torsion angles in Scheme 1, oscillations of the torsion angles α , γ and ζ occurs on a scale of tens ns ("10" in Scheme 2) while for the angle δ it was roughly 1 ns ("1" in Scheme 2). The motion of the two ribose units in the DMA molecules is not the same since the 3'- and 5'-end are not structurally equivalent. The faster oscillations were calculated for the 5'-end (f in Scheme 2). The oscillations of torsion angle β can be classified as well confined around its trans orientation and the corresponding distribution function is unimodal.

Scheme 2:

$$\begin{array}{ccccccccccc} 1 & 1 & 10 & 10 & konst. & 10 & f \\ \cdots & \delta & - & \varepsilon & - & \zeta & - & \alpha & - & \beta & - & \gamma & - & \delta & \cdots \end{array}$$

The motion of α , γ and ζ torsion angles is strongly correlated since the concerted flips between their typical values were actually calculated. Each of the three torsion angles has, in principle, a trimodal distribution, but only certain combinations of their amplitudes were calculated. Only two of these combinations were populated massively in the MD simulation, the α - γ - ζ triad with magnitudes 300° - 60° - 300° and 60° - 180° - 60° were populated during the 52% and 19% of simulation time (Table 2), respectively, where the first corresponds to the canonical A-RNA. The first triad corresponds to the canonical A-RNA backbone pattern. Other triads had a fast oscillatory character, and their overall residence time was practically negligible (Table 2).

The base stacking interaction contributes to the stability of NA molecules. The base stacking geometry was classified by the complex geometry criteria for mutual orientation of adjacent nitrogenous bases (for more details see Ref³⁸). Then the base-stacking ration was evaluated as the number of snapshots having a geometry corresponding to the stacked bases relative to the total number of snapshots accumulated along a given trajectory (Table 2).

Table 2: The observed conformations of the α - γ - ζ structural segment, their calculated population in MD time [%] and relevance of this triad to the stacking [%] calculated from time component.

ζ	α	γ	Population	Relevance to stack
300°	300°	60°	52	72
60°	60°	180°	19	68
60°	osc. 60° ... 180°	60°	11	3
60°	300°	60°	8	55
60°	60°	60°	6	18
300°	300°	180°	2	-
300°	60°	180°	1	-
300°	60°	60°	<1	-
300°	osc. 60° ... 180°	180°	<1	-
180°	300°	60°	<1	-
180°	60°	60°	<1	-

The α - γ - ζ segment is very important part of the RNA backbone since the magnitudes of the torsion angles are strongly coupled and the corresponding motions are therefore concerted. A correlation between geometry of the α - γ - ζ segment and the base stacking in the DMP RNA molecules calculated using the MD simulations showed strong coupling of the

300°-60°-300° and 60°-180°-60° segment pattern with high ratio of the base stacking geometry which was 72% and 68%, respectively. Others geometries of the α - γ - ζ segment which were also calculated do not showed such high correlation with bases in stack geometry (Table 2).

The calculated ratio of base stacking was correlated with the sequence of bases in the four DMP molecules and effect of five initial conformations on overall behavior during the individual 50 ns MD simulations was also evaluated. The calculated base-stacking ratios actually depended on both aspects. However, the effect of sequence on stacking of nucleobases in the four RNA molecules calculated in this study was large. The calculated base stacking ratios were 64-75%, 2-67%, 65-77%, and 24-75% for the ApA, ApC, CpA, and CpC RNA molecule, respectively. The ranges correspond to the effect of initial structure on calculated base stacking ratio. According to the MD calculations the adenosine at the 3' end of the DMP molecule stabilizes the stacking geometry regardless the kind of nucleobase at the 5' end. This conclusion is agreement with the MD study by Norberg and Nillson.⁶⁰ Each of the five MD simulations for one DMP molecules starting with different structural pattern had effect on overall behavior calculated during the 50 ns trajectory. Especially for the α - γ - ζ segment where the torsion angles exchanged their magnitudes on the time scale of tens ns the overall populations of the torsion angles depended on starting

The J-couplings assigned to the backbone and glycosidic torsion angles (Table 1) were calculated as dynamical averages using the distribution functions calculated from the MD trajectories for the torsion angles and the respective Karplus equations.^{12,15,44,59,61} In some cases when more parameterizations of the Karplus equation for one J-coupling are known we used the averaged Karplus equations. The dynamically averaged J-couplings correlated with the measured values* are in good overall agreement. For purpose of description of the deviation of the calculated values from the experiment can be grouped according to a J-coupling type. The absolute average deviation of the calculated J-coupling from the experiment value was 0.4, 0.6, 1.6, 1.0, and 1.1 Hz for the ²J(P-C), ³J(P-C), ³J(P-H), ³J(H-H), and ³J(C-H) couplings, respectively. The theoretical calculations of the J-couplings indicated that obtaining the complete structural and dynamical information for backbone and glycosidic torsion angles for highly flexible RNA molecules is possible.

* The DMP molecules were synthesized and NMR measurements were performed at Institute of Organic Chemistry and Biochemistry, v.v.i., Academy of Science of the Czech Republic

2.2 J-couplings Around the Glycosidic Torsion Angle in Guanosine

Magnitude of the glycosidic torsion angle χ (Figure 1) reflects the orientation of nucleobases with respect to sugar. This structural information is very important for determination of the global topologies of the NA. Dependence of the $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings on the glycosidic torsion angle was calculated in the guanosine (rG) and deoxy-guanosine (dG) molecules taking into account other structural effects like sugar pucker conformation including the effect of solvent. The 1J -couplings $^1J(\text{C1}'\text{-H1}')$, $^1J(\text{C2}'\text{-H2}')$, $^1J(\text{C2}'\text{-H2}''')$, and $^1J(\text{C8-H8})$ were also calculated in order to analyze their dependence on the glycosidic torsion.

The computational study of the six J-couplings was focused on the following dependencies and effects which may influence magnitude of the J-couplings: 1) dependence on the glycosidic torsion angle, 2) effect of B- and Z-DNA backbone conformation, 3) effect of the ribose/deoxy-ribose, 4) effect of the sugar pucker, 5) effect of solvent and 6) effect of the base pairing in non-WC RNA base pairs (the non-WC base pairs are discussed in the following section).

The $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings possess dominant dependence on the glycosidic torsion angle since the spin-spin coupling pathway mostly coincides with the torsion. This dominant dependence can be however, modulated by other structural parameters like the parameters mentioned above. Firstly we fitted the Karplus equations (Equation (3.4)) for the two J-couplings with the calculated data which covered the full rotational freedom of the torsion angle χ . The phase shift ϕ (Equation (3.4)) which is usually constrained¹² was also included in the fitting procedure as free parameter. In agreement with Munzarová et al.⁶¹, who firstly used the flexible phase shift for this assignment, we obtained the ϕ shift larger by $\sim 10^\circ$ (Tables 3 and 4) what significantly differs from its commonly accepted value of 60° .¹²

The parameterizations were done for new structural assignments of the $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ coupling. The $^3J(\text{C8-H1}')$ coupling was assigned to the χ' torsion ($\text{H1}'\text{-C1}'\text{-N9-C8}$) which actually corresponds to the spin-spin coupling pathway.[†] The torsion angle χ' corresponds exactly to the torsion angle χ only providing the exact planarity of the nucleobase

[†] We calculated the Karplus equation for the $^3J(\text{C8-H1}')$ coupling assigned to the torsion angle χ which includes the correction term for the deviations from planarity of the nitrogen N9 of nucleobase (The work is specified as “in preparation” in the list of publications.).

at nitrogen N9 (Figure 1). The $^3J(\text{C4-H1}')$ coupling was assigned to the χ torsion. The same assignments were actually used previously by Munzarová et al.⁶¹

Table 3: The calculated A , B , C parameters [Hz] and the phase ϕ [degree] of the Karplus equation for the $^3J(\text{C4-H1}')$ coupling.

Nucleoside	A	B	C	ϕ
B-DNA, C2'-endo, gas phase	3.68	1.87	0.49	70.44
B-DNA, C3'-endo, gas phase	3.66	1.84	0.41	71.11
B-DNA, C2'-endo, implicit water	3.82	1.99	0.60	69.71
B-DNA, C3'-endo, implicit water	3.82	1.95	0.53	70.56
Z-DNA, C2'-endo, gas phase	3.82	1.85	0.42	70.62
Z-DNA, C3'-endo, gas phase	3.61	1.78	0.45	68.26
A-RNA, C2'-endo, gas phase	3.50	1.73	0.51	70.09
A-RNA, C3'-endo, gas phase	3.86	1.83	0.41	71.68

Table 4: The calculated A , B , C parameters [Hz] and the phase ϕ [degree] of the Karplus equation for the $^3J(\text{C8-H1}')$ coupling.

Nucleoside	A	B	C	ϕ
B-DNA, C2'-endo, gas phase	4.44	-0.60	0.29	68.89
B-DNA, C3'-endo, gas phase	4.75	-1.09	0.18	69.90
B-DNA, C2'-endo, implicit water	4.57	-0.77	0.39	68.92
B-DNA, C3'-endo, implicit water	4.96	-1.30	0.29	69.31
Z-DNA, C2'-endo, gas phase	4.34	-0.71	0.31	67.58
Z-DNA, C3'-endo, gas phase	4.57	-1.07	0.37	67.40
A-RNA, C2'-endo, gas phase	4.34	-0.66	0.27	69.71
A-RNA, C3'-endo, gas phase	5.12	-1.22	0.14	73.06

The effect of sugar pucker on calculated dependence of the $^3J(\text{C8-H1}')$ coupling was larger than that obtained for the $^3J(\text{C4-H1}')$ couplings. The variations in magnitudes due to different sugar pucker were smaller than 0.4 and 0.8 Hz for the $^3J(\text{C4-H1}')$ and $^3J(\text{C8-H1}')$ coupling, respectively. The sugar pucker effect was dominant in both syn and anti region while it was practically negligible for the glycosidic torsions between the two regions. The same trend as was calculated for the B-DNA nucleoside was actually obtained for the Z-DNA nucleoside and remained the same even when the PCM model of water solvent was applied (see Figures 4 in Ref³⁹).

The application of the PCM solvent model led to the increase of both $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings ranging from 0.1 to 0.4 Hz, respectively. The effect of solvent led to the increase of the A and B coefficients in the new parameterizations of both Karplus curves (Tables 3 and 4).

The base pairing interactions were modeled with the WC/SE RNA base pairs. The WC/SE base pairs consist of nitrogenous base interacting non-covalently with other nucleoside in the close vicinity of its glycosidic bond, i.e via so called “Sugar Edge” (e.g. as in Figure 6). For the accurate structural interpretation of the J-couplings in the RNA molecules such effect may be important. However, the calculations indicated only negligible effect of the WC/SE base pairing on magnitude of the J-couplings which actually nicely fit the Karplus equations.

The calculated impact of the backbone geometry on the 3J -couplings was negligible, as well as the effect of the base pairing.

The mentioned structural and solvent effects were calculated also for the 1J -couplings and the sugar pucker effect was found dominant for the $^1J(C2'-H2')$ and $^1J(C2'-H2'2)$ couplings while the $^1J(C1'-H1')$ and $^1J(C8-H8)$ couplings depended dominantly on the glycosidic torsion angle. The dependences calculated for the 1J -couplings generally brought only additional information with regard to determination of the χ torsion angle. In this respect the $^3J(C8-H1')$ and $^3J(C4-H1')$ couplings represent the best assignment. However, the 1J -couplings represent another although less significant NMR structural descriptors for determining the glycosidic torsion.

2.3 Trans-hydrogen J-couplings in Watson-Crick/Sugar Edge base pairing

The Watson-Crick/Sugar Edge (WC/SE) base pairs represent important family of the non-covalent contacts in RNA since majority of the base pairs was actually observed in the solid state. The RNA base pairs were generally systemized into the twelve families of well-defined binding patterns by Leontis et al.⁶² The classification utilizes combinations of the three interaction edges in RNA nucleosides (Figure 5): the Watson-Crick edge (WC), the sugar edge (SE), and the Hoogsteen or CH edge defined for the purine or pyrimidine nucleoside, respectively. The RNA nucleosides can interact via any of the three edges and the interaction patterns can be either cis or trans. Example of such WC/SE base pair is shown in Figure 6.

The well known type of base pairing which is dominant in DNA is the WC/WC base pairing. However, in RNA molecules the non-Watson-Crick type of base pairing dominates.

The spectroscopic detection of such non-covalent intermolecular contacts would have large impact on the accuracy of structures determined with the NMR spectroscopy.

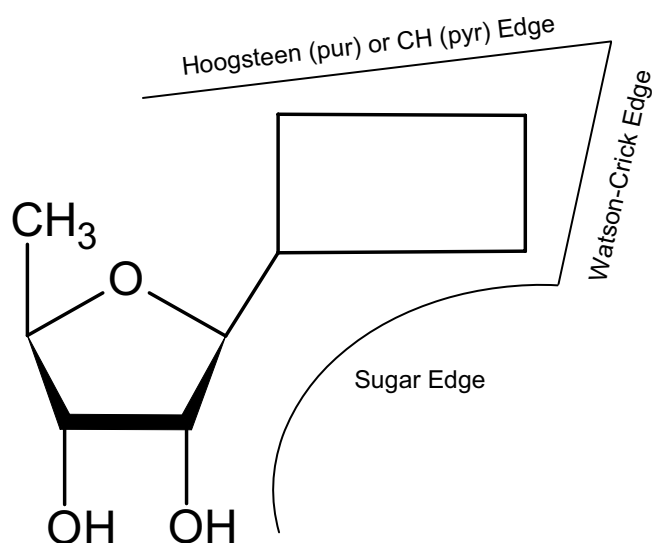


Figure 5: The interactions sides of RNA nucleoside.

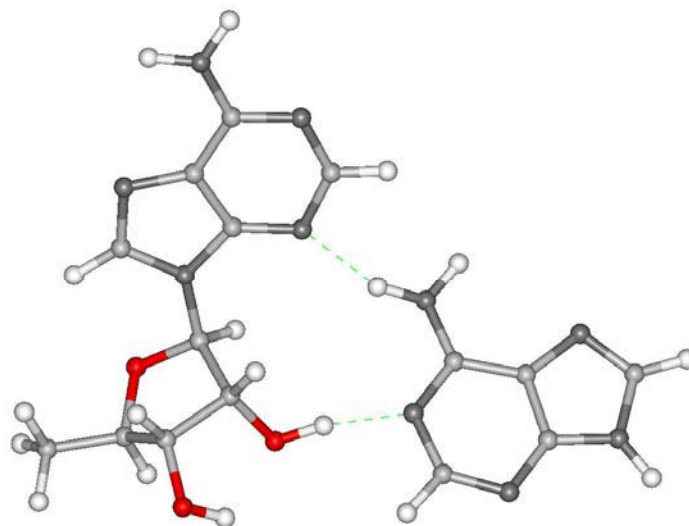


Figure 6: The *cis* A.rA RNA WC/SE base pair.

The study was carried out for all four RNA bases (Figure 1) where we used the following notation, X.rY (X, Y is A - adenine, C - cytosine, G - guanine, and U – uracil). X denotes the base and rY denotes the nucleoside interacting via their WC and SE edge, respectively.

The bases in the WC/SE base pairs interact via hydrogen bonds (H-bonds) and we therefore used in such case the ^hJ symbol for the trans-H-bond J-couplings. The non-covalent contacts in the WC/SE base pairs occur also between bases and sugar. All contacts were classified into different local bonding patterns including the four base-to-base and five sugar-to-base patterns.²¹ The calculated J-couplings were assigned to relevant geometry coordinates of these patterns.

The base-to-base WC/SE base pairs should be detectable with the ^hJ -couplings across the N-H \cdots N and N-H \cdots O=C links where the N-H group may belong both to the amino or imino group. The calculated $^{2h}\text{J}(\text{N-N})$, $^{1h}\text{J}(\text{N-H})$, $^{3h}\text{J}(\text{C-N})$, and $^{2h}\text{J}(\text{C-H})$ couplings generally decay with the separation of RNA bases and with the declination of link atoms from their collinear arrangement.

The $^{2h}\text{J}(\text{N-N})$ coupling depends on the nitrogen-nitrogen distance exponentially (Figure 1 in Ref²¹). Although similar trend was calculated for the $^{1h}\text{J}(\text{N-H})$ coupling the relatively large dispersion of its values in the WC/SE complexes diminishes its reliable correlation with any geometric parameter (Figure 2 in Ref²¹). Large dispersion of the calculated ^hJ couplings in WC/SE base pairs is caused by the deviations from planar arrangement of the bases. Both $^{2h}\text{J}(\text{N-N})$ and $^{1h}\text{J}(\text{N-H})$ couplings seemed to be almost independent of the kind of RNA base involved in the pairing.

The mutual correlation of the $^{3h}\text{J}(\text{C-N})$ and $^{2h}\text{J}(\text{C-H})$ couplings calculated across the N-H \cdots O=C bond show the linear dependence, but the slope of the calculated dependencies differs for the contacts involving the amino and imino group. The linear fit calculated for the correlation of the amino and imino contacts was $^{3h}\text{J} = 2.070 \cdot ^{2h}\text{J} + 0.420$ [Hz] and $^{3h}\text{J} = 1.186 \cdot ^{2h}\text{J} + 0.113$ [Hz] (Figure 7), respectively. This linear fit for imino contacts agrees with the experimental data measured for such imino contacts in proteins^{63,64} (Figure 4 in Ref²¹). Although the calculated different slope of the dependences can be hardly used for distinguishing between the amino and imino contacts, it indicates that the slope of the $^{3h}\text{J}(\text{C,N}) / ^{2h}\text{J}(\text{C,H})$ correlation calculated across the C=O \cdots H-N link is most probably rather conservative in the both RNA and proteins.

Several interactions may occur between the sugar hydroxyl group and atoms of the WC edge of RNA base and consequently many different types of the J-couplings can be correlated with particular bonding patterns. Calculated magnitudes of all ^hJ -couplings for the atoms separated by three and more bonds were smaller than 0.2 Hz. Only the $^{1h}\text{J}(\text{H-N})$ coupling between the hydrogen of sugar hydroxyl and nitrogen of RNA base had non-

negligible values. In comparison with the couplings calculated across the N-H...N link the $^1\text{J}(\text{H-N})$ couplings across the sugar-to base contacts were smaller.

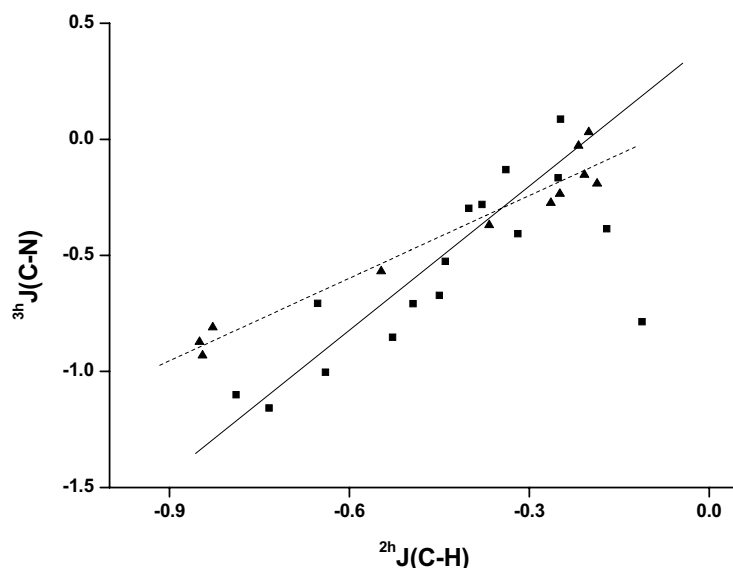


Figure 7: The calculated mutual correlation of the $^3\text{hJ}(\text{C-N})$ and $^2\text{hJ}(\text{C-H})$ couplings [Hz] across the N-H...O=C link involving amino (■) or imino (▲) group of the RNA base in the WC/SE RNA base pairs. — is the linear fit for amino, --- is the linear fit for imino contacts.

Intramolecular $^1\text{J}(\text{N-H})$ couplings in the amino and imino group involved in the base-to-base and sugar-to-base contacts possess large variation and structurally significant splitting for different bonding patterns. Their values ranges from -98 to - 81 Hz. The variation of the $^1\text{J}(\text{N-H})$ coupling in the amino group was larger in comparison with the couplings calculated in the imino group what correspond to the larger flexibility of the amino group. The calculated $^1\text{J}(\text{N-H})$ couplings indicate further the possibility for detection of the deflection from pyramidalization of the NH_2 group. The more pyramidal amino group with shorter N-H bond length have larger magnitude of the $^1\text{J}(\text{N-H})$ coupling.

Water molecule can penetrate inside the RNA base pairs influencing their structure and stability. For example, the cis-G.rG and trans-U.rG requires one water molecule between the bases for their better stability. The NMR calculations showed that insertion of water molecule inside the base pair damps the intercomplex ^hJ -couplings to zero.

The correlation of the calculated J-couplings with the molecular structure of WC/SE RNA base pairs as well as their good agreement with available NMR experiments indicate

large potential for their use in RNA structure validation. Structural variability of the WC/SE base pairs and frequent deviations of their geometries from ideal planar patterns result in effective damping and large dispersion of the calculated J-couplings which must be taken in the account in their accurate structural interpretation.

3 Conclusions

The first part of this thesis was focused on the theoretical modeling of the J-couplings in sugar-phosphate backbone in NA using the methods of quantum chemistry. Twenty different J-couplings between the ^{31}P , ^{13}C , and ^1H atoms were calculated including the solvent effect modeled with the implicit water solvent. The effect of specific explicit water molecules on NMR parameters was also studied. Different conformations of NA backbone were considered and the calculated J-couplings were assigned to the backbone torsion angles. For some backbone structural classes a unique combination of different J-couplings can provide the “fingerprint” which characterizes a particular sugar-phosphate conformation. The multidimensional character was calculated for some Karplus equations assigned to the backbone torsion angles which should improve their accuracy. The α and ζ torsion angles were newly assigned to ^2J -couplings allowing the discrimination between the gauche \pm and trans conformation of the α and ζ torsion angles. Good agreement of the calculated NMR parameters with available experiments was obtained.

The second part of the thesis concerns the dynamic behavior of the ApA, ApC, CpA and CpC RNA dinucleoside monophosphates studied using the MD simulations. The dynamical behavior of the RNA molecules was considered in the calculation of averaged J-couplings assigned to the backbone and glycosidic torsion angles. The theoretical results were compared with experimental J-couplings and the achieved overall good agreement thus validates different dynamical behavior of the four RNA molecules. The calculated dynamical behavior is different in different parts of the RNA molecules. Some torsion angles exhibit concerted motions like the α - ζ - γ triad and its 300° - 60° - 300° and 60° - 180° - 60° conformers belonging to the A-RNA class of RNA strongly correlate with maximal base stacking calculated with MD.

The third part of the thesis was focused on theoretical modeling of the J-couplings in close vicinity of the glycosidic torsion in guanosine. Two ^3J -couplings and four ^1J -couplings were calculated in dependence on the glycosidic torsion angle variation considering also other structural effects such as sugar pucker, backbone conformation, kind of sugar, including the solvent effect. New parameterization of the Karplus curves for both ^3J -couplings was done

including all mentioned structural effects. Sugar pucker have strong influence on the parameters of the Karplus equations. The two of the ^1J -couplings depend dominantly on sugar type (deoxy-ribose and ribose) and its pucker while the second two depend dominantly on the glycosidic torsion angle.

Last part of the thesis was focused on detection of the non-WC interactions in RNA, especially on the calculation of J-couplings across the H-bonds in cis and trans-WC/SE RNA base pairs. Several different bonding patterns were classified in the WC/SE base pairs and the J-couplings were assigned to the local bonding geometry parameters. Although the calculated data were not possible to assign only to the distance of RNA bases due to the dispersion caused by deviations from ideal planar arrangement of the bases in WC/SE base pairs, the calculated J-couplings show potential for detection of the inter-molecular non-covalent contacts in RNA with the NMR spectroscopy.

Used Instrumentation

Research was carried out in the computer facilities of the Department of Molecular Spectroscopy and Center for Biomolecules and Complex Molecular Systems at the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. They consist of

AMD Opteron Cluster, 2.6 GHz, 46 processors, LinuxOS, memory up to 4 GB

AMD Opteron Cluster, 2.6 GHz, 16 processors, LinuxOS, memory up to 2 GB

AMD Athlon XP, 1.8 GHz, 36 processors, LinuxOS, memory up to 1 GB

AMD Athlon XP, 3.2 GHz, 107 processors, LinuxOS, memory up to 1 GB

Intel Xeon Cluster, 2.4 GHz, 128 processors, LinuxOS, memory up to 4 GB

Intel Pentium 4, 2.8 GHz, 123 processors, LinuxOS, memory up to 1 GB

References

- (1) Noller, H. F. *Science* **2005**, 309, 1508.
- (2) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, 171, 737.
- (3) Sussman, J. L.; Seeman, N. C.; Kim, S.-H.; Berman, H. M. *J. Mol. Biol.* **1972**, 66, 403.
- (4) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
- (5) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, 31, 147.
- (6) Gilbert, W. *Nature* **1986**, 319, 618.
- (7) Gesteland, R. F., T.R. Cech, and J.F. Atkins,. *Cold Spring Harbor Monograph Series* **2006**.
- (8) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA World*, second ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1999; Vol. 37.
- (9) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys.J.* **1992**, 63, 751.
- (10) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, 28, 235.
- (11) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, 112, 535.
- (12) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. NMR Spectrosc.* **1998**, 32, 287.
- (13) Alkorta, I.; Elguero, J. *Int. J. Mol. Sci.* **2003**, 4, 64.
- (14) Englander, S. W.; Mayne, L. *Annual Review of Biophysics and Biomolecular Structure* **1992**, 21, 243.
- (15) Marino, J. P.; Schwalbe, H.; Griesinger, C. *Acc. Chem. Res.* **1999**, 32, 614.
- (16) Schweitzer, B. A.; Kool, E. T. *J. Org. Chem.* **1994**, 59, 7238.
- (17) Židek, L.; Štefl, R.; Sklenář, V. *Curr. Opin. Struct. Biol.* **2001**, 11, 275.
- (18) Qin, P. Z.; Dieckmann, T. *Curr. Op. Struct. Biol.* **2004**, 14, 350.
- (19) Grzesiek, S.; Barfield, M.; Cordier, F.; Dingley, A.; Feigon, J.; Nicholson, L. K. *Abstracts of Papers of the American Chemical Society* **2000**, 219, U139.
- (20) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. *Prog. NMR Spectrosc.* **2004**, 45, 275.
- (21) Vokáčová, Z.; Šponer, J.; Šponer, J. E.; Sychrovský, V. *J. Phys. Chem. B* **2007**, 111, 10813.
- (22) Sychrovský, V.; Buděšínský, M.; Benda, L.; Špirko, V.; Vokáčová, Z.; Šebestík, J.; Bouř, P. *J. Phys. Chem. B* **2008**, 112, 1796.
- (23) Sychrovský, V.; Vokáčová, Z.; Šponer, J.; Špačková, N.; Schneider, B. *J. Phys. Chem. B* **2006**, 110, 22894.
- (24) Kaupp, M.; Malkin, V. G. *J. Comp. Chem.* **1999**, 20, V.
- (25) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, 99, 293.
- (26) Szabo, A.; Ostlund, N. S. *McGraw-Hill Publishing Company* **1982**.
- (27) Skála, L. *Charles Univerzity, Karolinum* **1994**.
- (28) Widmark, P. E. *Lund University* **2007**.
- (29) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136, B864.

- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A. Gaussian 03, Revision C.02. In *Gaussian, Inc.* Wallingford, 2004.
- (31) Case, D. A., Darden, T.A., Cheatham, III, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Merz, K.M., Wang, B., Pearlman, D.A., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J.W., Ross, W.S. and Kollman, P.A. . AMBER 8 University of California, San Francisco, 2004.
- (32) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 1049.
- (33) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.
- (34) Sychrovský, V.; Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2000**, *113*, 3530.
- (35) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.
- (36) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR - Basis Principles and Progress*; Springer: Heidelberg, 1990.
- (37) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **2000**, *104*, 5631.
- (38) Vokáčová, Z.; Buděšínský, M.; Rosenberg, I.; Schneider, B.; Šponer, J.; Sychrovský, V. *J. Phys. Chem. B* **2009**, *113*, 1182.
- (39) Vokáčová, Z.; Bickelhaupt, F. M.; Šponer, J.; Sychrovský, V. *J. Phys. Chem. A* **2009**, *accepted*.
- (40) Brumovská, E.; Sychrovský, V.; Vokáčová, Z.; Šponer, J.; Schneider, B.; Trantírek, L. *Journal of Biomolecular Nmr* **2008**, *42*, 209.
- (41) Vokáčová, Z.; Burda, J. V. *Journal of Physical Chemistry A* **2007**, *111*, 5864.
- (42) Schneider, B.; Morávek, Z.; Berman, H. M. *Nucleic Acids Res.* **2004**, *32*, 1666.
- (43) Murray, L. J.; Arendall 3rd, W. B.; Richardson, D. C.; Richardson, J. S. *Proc.Natl.Acad.Sci.USA* **2003**, *100*, 13904.
- (44) Yokoyama, S.; Inagaki, F.; Miyazawa, T. *Biochemistry* **1981**, *20*, 2981.
- (45) Barbic, A.; Zimmer, D. P.; Crothers, D. M. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 2369.
- (46) Hines, J. V.; Landry, S. M.; Varani, G.; Tinoco, I. *J. Am. Chem. Soc.* **1994**, *116*, 5823.
- (47) Mooren, M. M. W.; Wijmenga, S. S.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Nucleic Acids Research* **1994**, *22*, 2658.
- (48) Sich, C.; Ohlenschlager, O.; Ramachandran, R.; Gorlach, M.; Brown, L. R. *Biochemistry* **1997**, *36*, 13989.
- (49) Richter, C.; Reif, B.; Worner, K.; Quant, S.; Marino, J. P.; Engels, J. W.; Griesinger, C.; Schwalbe, H. *Journal of Biomolecular Nmr* **1998**, *12*, 223.
- (50) Blommers, M. J. J.; Haasnoot, C. A. G.; Walters, J.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Biochemistry* **1988**, *27*, 8361.
- (51) Varani, G.; Cheong, C. J.; Tinoco, I. *Biochemistry* **1991**, *30*, 3280.

- (52) Zimmer, D. P.; Marino, J. P.; Griesinger, C. *Magn. Reson. Chem.* **1996**, *34*, S177.
- (53) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona, C. *Journal of Biomolecular Structure & Dynamics* **1984**, *1*, 1387.
- (54) Lapper, R. D.; Mantsch, H. H.; Smith, I. C. P. *J Am Chem Soc* **1973**, *95*, 2878.
- (55) Clore, G. M.; Murphy, E. C.; Gronenborn, A. M.; Bax, A. *Journal of Magnetic Resonance* **1998**, *134*, 164.
- (56) Sklenář, V.; Bax, A. *Journal of the American Chemical Society* **1987**, *109*, 7525.
- (57) Wu, Z. R.; Tjandra, N.; Bax, A. *Journal of Biomolecular Nmr* **2001**, *19*, 367.
- (58) O'Neil-Cabello, E.; Wu, Z. R.; Bryce, D. L.; Nikonowicz, E. P.; Bax, A. *Journal of Biomolecular Nmr* **2004**, *30*, 61.
- (59) Haasnoot, C. A. G.; Deleeuw, F.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
- (60) Norberg, J.; Nilsson, L. *J. Am. Chem. Soc.* **1995**, *117*, 10832.
- (61) Munzarová, M. L.; Sklenář, V. *J. Am. Chem. Soc.* **2003**, *125*, 3649.
- (62) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497.
- (63) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 1601.
- (64) Cordier, F.; Rogowski, M.; Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1999**, *140*, 510.

Other sources (in alphabetic order):

- Biochemie, Donald Voet a Judith G. Voetová, Victoria Publishing, 1995
- Experimentální metody biofyziky, Václav Prosser a kol., Academia, Praha 1989
- Pokročilé kapitoly z molekulární spektroskopie a teorie elektromagnetického záření, Petr Bouř, unpublished lecture notes, Praha, 2006
- Wikipedia, The Free Encyclopedia, <http://en.wikipedia.org/>

List of Publications

Present work is based on five papers listed below:

Calculation of structural behavior of indirect NMR spin-spin couplings in the backbone of nucleic acids

Sychrovský, V; Vokáčová, Z; Šponer, J; Špačková, N; Schneider, B
Journal of Physical Chemistry B, Volume: 110, Issue: 45, Pages: 22894-22902, 2006
Times Cited: 3

Theoretical study of the scalar coupling constants across the noncovalent contacts in RNA base pairs: The cis- and trans-Watson-Crick/sugar edge base pair family

Vokáčová, Z; Šponer, J; Šponer, JE; Sychrovský, V
Journal of Physical Chemistry B, Volume: 111, Issue: 36, Pages: 10813-10824, 2007
Times Cited: 2

Structure and Dynamics of the ApA, ApC, CpA, and CpC RNA Dinucleoside Monophosphates Resolved with NMR Scalar Spin-Spin Couplings

Vokáčová, Z; Buděšínský, M; Rosenberg, I; Schneider, B; Šponer, J; Sychrovský, V
Journal of Physical Chemistry B, Volume: 113, Issue: 4, Pages: 1182-1191, 2009
Times Cited: 0

Structural Interpretation of J-coupling Constants in Guanosine and Deoxy-Guanosine: Modeling the Effects of Sugar Pucker, Backbone Conformation and Base Pairing

Vokáčová, Z; Bickelhaupt, FM; Šponer, J; Sychrovský, V
Journal of Physical Chemistry A, accepted

Influence of the structural non-rigidity and nonplanarity of guanosine on determination of the glycosidic bond angle from $3J_{C4-H1'}$ and $3J_{C8-H1'}$ scalar couplings.

Vokáčová, Z; Trantírek L. and Sychrovský V.
in preparing, is not included in the section Attached Publication

Vokáčová, Z. also Publisher her results in Proceedings of Contributed Papers

They are not included in the part Attached Publication because published data are mostly included in authoress' publication listed above.

Quantum Chemistry Study of NMR Parameters of cis Watson-Crick/Sugar Edge RNA Base Pair Family

Vokáčová, Z

WDS'06 Proceedings of Contributed Papers: Part III – Physics, Prague, Matfyzpress, Pages: 64-69, 2006

The Molecular Dynamic Study of Stacking Interaction in RNA Dinucleosides, the Dependence of Stacking Ability on RNA Sequence

Vokáčová, Z

WDS'07 Proceedings of Contributed Papers: Part III – Physics, Prague, Matfyzpress, Pages: 78-83, 2007

Other publications of authoress:

Computational study on spectral properties of the selected pigments from various photosystems: Structure-transition energy relationship

Vokáčová, Z; Burda, J

Journal of Physical Chemistry A, Volume: 111, Issue: 26, Pages: 5864-5878, 2007

Times Cited: 3

Dependence of the L-alanyl-L-alanine conformation on molecular charge determined from Ab initio computations and NMR spectra

Sychrovský, V; Buděšínský, M; Benda, L; Špirko, V; Vokáčová, Z; Šebestík, J; Bouř, P

Journal of Physical Chemistry B, Volume: 112, Issue: 6, Pages: 1796-1805, 2008

Times Cited: 3

Effect of local sugar and base geometry on C-13 and N-15 magnetic shielding anisotropy in DNA nucleosides

Brumovská, E; Sychrovský, V; Vokáčová, Z; Šponer, J; Schneider, B; Trantírek, L

Journal of Biomolecular Volume: 42, Issue: 3, Pages: 209-223, 2008

Times Cited: 0

Number of citation is related to 30.6.2009

Attached Publications

Calculation of Structural Behavior of Indirect NMR Spin–Spin Couplings in the Backbone of Nucleic Acids

Vladimír Sychrovský,^{*,†,‡} Zuzana Vokáčová,[†] Jiří Šponer,[‡] Nad'a Špačková,[‡] and Bohdan Schneider^{*,†}

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Fleming Square 2, CZ-16610 Prague, Czech Republic, and Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Received: August 3, 2006

Calculated indirect NMR spin–spin coupling constants (J-couplings) between ^{31}P , ^{13}C , and ^1H nuclei were related to the backbone torsion angles of nucleic acids (NAs), and it was shown that J-couplings can facilitate accurate and reliable structural interpretation of NMR measurements and help to discriminate between their distinct conformational classes. A proposed stepwise procedure suggests assignment of the J-couplings to torsion angles from the sugar part to the phosphodiester link. Some J-couplings show multidimensional dependence on torsion angles, the most prominent of which is the effect of the sugar pucker. J-couplings were calculated in 16 distinct nucleic acid conformations, two principal double-helical DNAs, B- and A-, the main RNA form, A-RNA, as well as in 13 other RNA conformations. High-level quantum mechanics calculations used a baseless dinucleoside phosphate as a molecular model, and the effect of solvent was included. The predicted J-couplings correlate reliably with available experimental data from the literature.

Introduction

Functional diversity of nucleic acids (NAs) makes their study crucial for deeper understanding of biological processes. Structural studies of NAs have contributed to this understanding significantly since their advent in the 1970s.¹ The number of three-dimensional structures of nucleic acids determined by both major methods, X-ray crystallography and solution nuclear magnetic resonance (NMR), grows at a steadfast pace due to the progress of both of these methods, and the available structures² make it more and more evident that functional variability of NAs is reflected by their structural diversity.³

A variety of advanced NMR techniques can be used for determination of structures of NAs. These methods make use of diverse physical effects, including the nuclear Overhauser effect (NOE), the NMR shift, the direct as well as indirect spin–spin coupling, and the cross-correlation relaxation rates.^{4–8} All of these types of experiments can provide valuable information about the conformation of NAs. Information solely from the internuclear distance restraints derived from NOEs does not seem to be sufficient to determine DNA or RNA structures at high resolution because of the low density of proton nuclei with a measurable signal along the backbone.⁶ Interpretation of NMR shifts in terms of the local molecular structure is also not usually straightforward even when recent applications correlating the cross-correlation relaxation rates between the chemical shift anisotropy (CSA) and the relevant bond vector are promising.⁹ These experiments may be able to determine the relative orientation between the CSA and the bond vector and therefore may be helpful in evaluation of the local geometry. The measurement of the indirect spin–spin coupling constants, “J-

couplings”, provides independent, and often quite specific, information about the conformations between the coupled nuclei.^{10,11}

Despite these advanced NMR techniques, the number of measured NMR parameters and derived geometric constraints available for a structure analysis is still rather limited. What might also hamper our ability to determine high-resolution NA structures is a lack of rules and tools to interpret already measured parameters. These interpretation rules must respect the complicated nature of NA conformational space. In this work, we take advantage of a priori knowledge of populated NA conformations for structural interpretation of J-couplings between ^{31}P , ^{13}C , and ^1H nuclei of the NA backbone.

Calculations by modern quantum mechanics methods can complement structural interpretation of measured NMR parameters.^{12,13} In this work, J-couplings were calculated by the state-of-the-art quantum mechanics methods for 16 representative conformations of an NA fragment, dinucleoside phosphate (n–P–n). Selection of conformations of the calculated n–P–n fragments was the key initial step of the analysis; first and foremost, biologically important and well-characterized double-helical forms, B-DNA, A-DNA, and A-RNA, were calculated. A proper representation of structurally extremely variable non-A-RNA was a more challenging task. The RNA backbone has been shown to be classifiable into a relatively small number of classes,^{14,15} and these classes were represented by 13 examples in our calculations.

The analysis of the calculations presented in this work suggests that the J-couplings can contribute to determination of the NA backbone. On the basis of our calculations, we suggest that the J-couplings can define conformational regions of most backbone torsion angles and thus facilitate structural interpretation of NMR measurements. In some cases, a unique combination of a few spin–spin couplings provides a “fingerprint” that characterizes a particular n–P–n conformation.

* To whom the correspondence should be addressed. E-mail: sychrovsky@uochb.cas.cz (V.S.); bohdan@uochb.cas.cz (B.S.).

[†] Institute of Organic Chemistry and Biochemistry.

[‡] Institute of Biophysics.

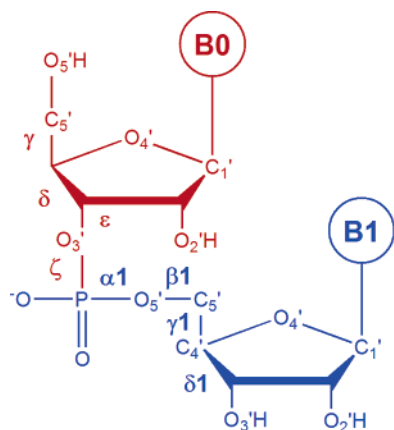


Figure 1. Dinucleoside monophosphate (n–P–n). The first nucleoside is shown in red, the second in blue. In the test calculations, B0 was adenine, and B1 was uracil base (n–P–n is adenosyl-3′-5′-uridine phosphate); in the calculations, both B0 and B1 were replaced by methyl groups. Both 5′- and 3′-ends are terminated by hydroxyl groups. In text and tables, all corresponding atoms and torsions are labeled while extension “1” was used to mark the second nucleoside (blue). In the figure, the key atoms are labeled, and the backbone torsion angles are defined in the standard way: $\gamma = \text{O5′-C5′-C4′-C3′}$, $\delta = \text{C5′-C4′-C3′-O3′}$, $\epsilon = \text{C4′-C3′-O3′-P}$, $\zeta = \text{C3′-O3′-P-O5′1}$, $\alpha1 = \text{O3′-P-O5′1-C5′1}$, $\beta1 = \text{P-O5′1-C5′1-C4′1}$.

Methods

Selection of Model Compounds. Dinucleoside 3′-5′-monophosphate, n–P–n, shown in Figure 1, represents the smallest molecule that has the key chemical and structural features of nucleic acids yet is accessible by quantum mechanics calculations. A chemically and conformationally key phosphodiester link is in the center of the n–P–n molecule and connects the phosphorus nuclei to the other atoms of the backbone. Because the nature of NMR properties is local, the n–P–n molecule is a relevant molecular model for theoretical investigation of J-couplings in the vicinity of the phosphate group.

Selection of Dinucleoside Monophosphate (n–P–n) Conformations. The backbone of n–P–n is described by seven torsion angles: δ , ϵ , ζ , $\alpha1$, $\beta1$, $\gamma1$, and $\delta1$; their actual values for the studied n–P–n conformational classes are shown in Table 1.

n–P–n conformations were selected that represent the most important conformational classes of nucleic acids: the most important DNA conformations, A and BI forms, the canonical A-RNA, and 13 other RNA conformations, including two minor A-RNA and 11 non-A conformations. Conformationally less usual conformations (families 22, 19, 1, 2, 3, 4, 6, 7, 9, 12, 15, 16, and 18 in Table 1) were selected in order to cover populated regions of the backbone torsion angles and their combinations most typical outside of the A and B classes; these distinct conformations were taken from a previous study¹⁵ and are labeled hereafter by numbers used in Table 3 of this reference.

Calculation Methods. Initial models of dinucleoside phosphates were constructed as reported previously for RNA fragments¹⁵ and for the most prevalent DNA conformations, BI and A.¹⁶ These “experimental” geometries were relaxed using the optimization procedure with the molecular mechanics program Amber.¹⁷ Torsion angles of the backbone were restrained in order to keep the overall conformation of the studied classes close to their original values; the values before and after the relaxation procedure differ by less than 10°. The largest deviations were observed mostly for α and γ torsions, but the deviations tend to cancel each other. After geometry relaxation, the nitrogenous bases were substituted by the methyl

TABLE 1: Values of Torsion Angles (deg) along the Backbone in n–P–n Dinucleoside Monophosphates Used for the NMR Calculations

conformation ^a	δ	ϵ	ζ	$\alpha1$	$\beta1$	$\gamma1$	$\delta1$
BI-DNA	130	184	271	303	180	39	139
A-DNA	84	212	285	294	179	59	87
A-RNA (“canonical”, 20)	84	212	285	294	179	59	87
A-RNA (clone 22)	150	266	294	295	191	62	87
A-RNA (clone 19)	82	196	289	157	196	181	86
1	79	203	56	164	152	50	144
2	83	198	286	162	163	54	83
3	85	235	205	296	153	52	87
4	148	270	81	70	188	176	80
6	144	212	163	293	167	49	144
7	144	192	156	289	149	49	87
9	148	234	83	74	172	57	144
12	79	214	165	289	166	51	88
15	82	221	210	69	195	48	85
16	83	224	208	66	196	47	143
18	85	256	260	63	179	301	145

^a RNA conformations are numbered according to the numbering in Table 3 of ref 15.

group and both the 5′ and 3′ ends were terminated by hydroxyl groups (Figure 1).

The coupled perturbed DFT method^{13,18,19} with the B3LYP DFT functional was used for calculation of the indirect NMR spin–spin coupling constants by including all four coupling terms: the diamagnetic spin–orbit (DSO), paramagnetic spin–orbit (PSO), Fermi–contact (FC), and spin–dipolar (SD). The J-couplings were calculated for isotopes ¹H, ¹³C, and ³¹P; the index referring to the isotope identity is skipped in the following text for clarity. The atomic basis, usually called Iglo II and Iglo III,²⁰ was used in the NMR calculations. All calculations were done with the Gaussian 03 program package.²¹

Solvent Effect on Calculated J-Couplings. The solvent is inherently present during the NMR spectra measurement and its role in theoretical calculations of NMR parameters thus needs to be analyzed.^{22–24} To validate the use of the polarized continuum solvent (PCM) model²⁵ in actual NMR calculations, we tested the performance of several solvent models: (a) in vacuo, (b) the PCM, (c) explicit hydration of the phosphate, and (d) the explicit hydration as in (c) plus the PCM solvent.

The B-DNA form has a well-characterized first hydration shell and was therefore used to test the effect of hydration on calculated J-couplings. An empirical study²⁶ has shown that each phosphate is hydrated by six water molecules and the positions of oxygen atoms of these water molecules are known;²⁶ to determine plausible hydrogen coordinates, the complex was gradient optimized at the B3LYP level with the 6-31G** atomic basis while keeping the backbone torsion angles as well as the angular orientation of the water molecules fixed. The J-couplings shown in Table 2 were calculated for the n–P–n with nitrogenous bases substituted by methyl groups. Valence geometry of the n–P–n has a notable effect on both ²J(P,C) couplings shown in Table 2, and the in vacuo and PCM J-couplings obtained for geometries optimized by AMBER and B3LYP generally differ by less than 1.6 Hz. The trend caused by PCM hydration is the same in both hydration models of B-DNA and the shift in J values caused by hydration was for all calculated J-couplings smaller than 2.2 Hz. The performance of explicit hydration is fully comparable with the PCM one for all calculated J-couplings except the cases of ³J(P,C4′) and ³J-(P,C4′1) coupling when the effect of explicit hydration increases twice. This fact validates good overall performance of the PCM hydration model used for actual calculations of J-couplings. Further hydration by PCM of the B-DNA complex with

TABLE 2: Spin–Spin Coupling Constants^a [Hz] Calculated in B-DNA with Different Models of Hydration

J-coupling	hydration model					
	in vacuo ^b	in vacuo ^b PCM	in vacuo ^c	in vacuo ^c PCM	explicit hydration ^c	explicit hydration ^c PCM
² J(P,C3')	−7.7	−6.8	−9.2	−8.0	−8.1	−8.0
² J(P,C5'1)	−8.3	−8.1	−9.5	−9.0	−8.8	−9.1
² J(C5'1,C3'1)	0.8	0.8	0.1	0.1	0.1	0.1
² J(C5'1,H4'1)	4.3	5.7	5.5	6.6	7.4	7.4
² J(C4'1,1H5'1)	−4.5	−4.7	−4.4	−4.6	−4.6	−4.6
² J(C4'1,2H5'1)	4.1	2.8	4.6	3.0	3.0	2.5
³ J(P,C4'1)	10.0	11.2	10.5	11.9	13.5	13.5
³ J(P,C4')	11.0	12.3	12.1	13.9	15.8	16.0
³ J(P,C2')	0.5	0.4	0.6	0.4	0.3	0.3
³ J(P,H3')	2.4	2.5	2.3	2.5	2.9	2.9
³ J(P,1H5'1)	2.7	2.7	3.5	3.4	3.3	3.2
³ J(P,2H5'1)	1.8	2.0	1.1	1.1	1.2	1.2
³ J(C5'1,H3'1)	2.7	3.0	2.9	3.1	3.1	3.2
³ J(1H5'1,H4'1)	0.9	0.9	1.4	1.4	1.4	1.4
³ J(2H5'1,H4'1)	4.1	4.0	2.8	2.6	2.7	2.6
³ J(H2',H3')	6.8	6.8	5.8	5.7	5.7	5.6
³ J(2H2',H3')	0.6	0.5	0.4	0.3	0.4	0.4
³ J(H4',H3')	1.6	1.5	1.3	1.2	1.2	1.2
³ J(H4'1,H3'1)	0.6	0.6	1.1	1.2	1.2	1.2
⁴ J(P,C3'1)	−0.9	−0.8	−0.6	−0.4	−0.4	−0.4
⁴ J(P,H4'1)	2.0	3.5	2.9	5.1	5.9	6.8

^a The basis set for NMR calculation was Iglo II. ^b NA geometry optimization with AMBER (parm99 force field) method was done without explicit hydration. ^c NA geometry optimization with DFT B3LYP method was done including explicit hydration. The pattern of explicit waters was taken from a study of phosphate hydration in double-helical DNA;²⁶ for coordinates, see the Supporting Information.

explicitly hydrated phosphate brings only small change in magnitude of calculated J-couplings, mostly smaller than 0.5 Hz, with the limit value of 0.9 Hz obtained for the ⁴J(P,H4'1) coupling.

The PCM hydration accounts for the dominant part of the change in calculated J-couplings between in vacuo and the hydrated models. In addition, magnitudes of J-couplings calculated with the PCM model and by explicit water molecules are different, but their deviations follow the same trend, in agreement with our previous finding,²⁴ and this trend is almost independent of the local molecular geometry. Further hydration of the explicitly hydrated phosphate group by PCM brings about a small change of the calculated J-couplings. The PCM method thus represents the optimal compromise between the tested hydration models. Limits of the PCM model are, however, obvious, and the model should be used with care, and predictions for J-couplings should be cross-validated in the context of the local geometry.

Effect of Substitution of Nitrogenous Bases and the Effect of Atomic Basis Set. Although all NMR properties are local by nature, we performed a number of test calculations to estimate the influence of nitrogenous bases on backbone J-couplings and to validate their replacement by methyl groups. The test was done for adenosyl-3'-5'-uridine phosphate in the canonical A-RNA conformation. The J-couplings shown in Table 3 were calculated for n–P–n with and without bases in the PCM solvent and in vacuo using basis sets Iglo II and Iglo III. Substituting the A, U bases to methyl groups changes the calculated couplings by less than 0.5 Hz so that their effect is smaller than the effect of hydration.

The effect of size of the atomic basis set on J-couplings was estimated for the model with PCM hydration and n–P–n with nitrogenous bases substituted by methyl groups; the differences between Iglo II and Iglo III calculations are smaller than 0.6 Hz for all couplings, and the effect is comparable with substituting nitrogenous bases by methyl groups and is smaller than the effect of hydration, as shown in Table 3 (last two columns, in vacuo calculations).

TABLE 3: Effect of Substitution of Nitrogenous Bases in Adenosyl-3'-5'-uridine Phosphate by Methyl Groups on the J-couplings [Hz] Calculated in the Canonical A-RNA^a

J-coupling	PCM Iglo III (−CH ₃) ₂	PCM Iglo II A, U bases	PCM Iglo II (−CH ₃) ₂	in vacuo Iglo II A, U bases	in vacuo Iglo III (−CH ₃) ₂
² J(P,C3')	−5.7	−6.2	−6.0	−7.2	−6.4
² J(P,C5'1)	−6.9	−7.1	−7.3	−7.9	−7.3
² J(C5'1,C3'1)	−0.6	−0.7	−0.8	−0.7	−0.6
² J(C5'1,H4'1)	7.7	6.6	7.1	5.5	6.1
² J(C4'1,1H5'1)	−4.4	−5.0	−4.9	−4.7	−4.1
² J(C4'1,2H5'1)	3.4	3.4	3.0	4.3	4.5
³ J(P,C4'1)	11.3	11.2	11.5	10.1	10.4
³ J(P,C4')	8.8	8.6	8.8	7.7	7.9
³ J(P,C2')	0.2	0.2	0.2	0.1	0.1
³ J(P,H3')	8.8	8.8	8.9	8.1	8.0
³ J(P,1H5'1)	2.0	2.3	2.3	2.4	2.1
³ J(P,2H5'1)	1.5	1.8	1.7	1.6	1.3
³ J(C5'1,H3'1)	4.3	4.4	4.3	4.0	3.9
³ J(1H5'1,H4'1)	2.4	2.6	2.6	2.7	2.4
³ J(2H5'1,H4'1)	1.1	1.4	1.4	1.4	1.2
³ J(H2',H3')	4.7	5.0	4.9	5.1	4.7
³ J(H4',H3')	8.8	8.3	8.4	8.2	8.8
³ J(H4'1,H3'1)	8.6	8.1	8.1	7.8	8.3
⁴ J(P,C3'1)	−0.5	−0.6	−0.5	−0.7	−0.6
⁴ J(P,H4'1)	3.7	3.2	3.4	2.0	2.3

^a Geometry optimized with Amber.

Results

Calculated indirect coupling constants are grouped according to the type of nuclei involved in spin–spin interaction; the J-couplings involving a phosphorus atom ⁿJ(P,X), where X = C, H are shown in Table 4, while the couplings involving carbon and hydrogen, ⁿJ(X,Y) (X,Y = C, H), are shown in Table 5.

The calculated J-couplings are presented in the following text in context with backbone torsion angle(s) with which they are primarily correlated. Selected data for torsion angles along the backbone in n–P–n dinucleosides are shown in Figure 2, together with the experimental J-couplings and Karplus curves from the literature. The experimental couplings in Figure 2 were correlated with the information about torsion angles; this

TABLE 4: Calculated Indirect Spin–Spin Coupling Constants [Hz] Involving Phosphorus Atom ${}^2J(P, X)$, X Is C or H

conformation ^a	${}^2J(P, C)$		${}^3J(P, C)$			${}^3J(P, H)$			${}^4J(P, C)$	${}^4J(P, H)$
	C3' ζ^b	C5' α^b	C4' β^b	C4' ϵ^b	C2' ϵ^b	H3' ϵ^b	1H5' β^b	2H5' β^b	C3' γ^b	H4' γ^b
B-DNA	−6.4	−7.7	11.1	12.3	0.4	2.4	2.5	1.7	−0.8	3.8
A-DNA	−5.9	−7.0	11.1	9.3	0.2	8.7	2.0	1.5	−0.5	3.7
A-RNA, canon	−5.7	−6.9	11.3	8.8	0.2	8.8	2.0	1.5	−0.5	3.7
A-RNA, 22	−5.4	−7.1	11.2	−0.2	7.5	8.1	0.3	4.0	−0.5	3.5
A-RNA, 19	−6.1	−5.1	10.7	12.3	−0.1	5.3	−0.4	4.3	1.1	−1.0
1	−6.5	−5.3	7.8	10.7	−0.1	7.6	6.5	−0.2	−1.0	4.3
2	−6.1	−4.4	9.6	11.9	−0.2	5.9	4.3	−0.5	−0.5	3.8
3	−3.4	−7.2	8.9	3.9	2.8	10.1	8.2	0.2	−0.3	3.8
4	−6.6	−7.1	11.3	0.0	7.7	6.3	0.7	3.2	0.9	−1.1
6	−3.5	−7.5	10.4	8.7	−0.1	8.3	4.8	0.0	−0.8	4.6
7	−3.8	−6.1	8.6	11.0	0.3	4.4	9.3	0.7	−0.2	3.8
9	−6.3	−7.3	10.7	3.4	2.7	11.5	3.3	0.8	−0.9	3.9
12	−3.0	−7.3	10.7	8.7	0.3	8.6	4.9	−0.1	−0.2	4.0
15	−3.0	−7.0	11.2	7.0	1.2	9.6	−0.2	6.1	−0.4	3.5
16	−3.1	−7.3	10.9	6.3	1.5	10.0	−0.2	6.7	−0.7	3.8
18	−5.5	−7.1	13.6	0.7	6.1	9.6	1.5	2.3	−0.7	−1.2

^a Torsion angles are listed in Table 1. ^b The J-couplings are discussed primarily in relation to the indicated torsion angles.

TABLE 5: Calculated Indirect Spin–Spin Coupling Constants^b in Hz Involving Carbon and Hydrogen Atoms ${}^2J(C, C)$, ${}^2J(C, H)$ and ${}^3J(H, H)$

conformation ^a	² J(C,C)		² J(C,H)		³ J(C,H)	³ J(H,H)				
	C5'1	C5'1	C4'1	C4'1	C5'1	H4'1	H4'1	H2'	H3'	H3'1
	C3'1	H4'1	1H5'1	2H5'1	H3'1	1H5'1	2H5'1	H3'	H4'	H4'1
	$\gamma 1^b$	$\gamma 1^b$	$\gamma 1^b$	$\gamma 1^b$	$\gamma 1^b$	$\gamma 1^b$	$\gamma 1^b$	δ^b	δ^b	$\delta 1^b$
B-DNA	0.9	6.3	−4.4	3.2	2.8	0.7	3.5	6.5	1.4	0.5
A-DNA	−0.7	7.9	−4.4	3.5	4.0	2.3	1.1	7.1	8.4	8.1
A-RNA, 20	−0.6	7.7	−4.4	3.4	4.3	2.4	1.1	4.7	8.8	8.6
A-RNA, 22	−0.6	7.5	−4.0	2.7	4.3	2.6	1.1	5.2	0.1	9.1
A-RNA, 19	4.9	−2.7	3.6	−5.0	5.5	2.1	9.5	4.9	9.2	8.9
1	1.3	7.8	−4.1	3.9	2.4	1.8	1.5	4.3	9.3	0.2
2	−1.0	8.8	−4.5	3.7	4.2	1.9	1.4	4.8	9.2	9.0
3	−0.8	7.4	−4.7	4.5	3.9	1.8	1.4	4.9	9.1	8.5
4	3.9	−1.5	3.2	−4.5	5.4	2.4	9.5	5.4	0.2	9.5
6	0.9	7.6	−4.2	3.4	2.6	1.6	1.8	5.1	0.2	0.2
7	−0.9	7.6	−4.8	4.5	4.1	1.5	1.7	4.7	0.3	8.6
9	1.4	7.2	−3.8	3.1	2.6	2.6	1.1	5.1	0.1	0.1
12	−0.8	7.8	−4.6	4.0	4.2	1.6	1.2	4.5	9.2	8.6
15	−0.9	7.7	−3.9	2.6	4.1	1.1	2.3	4.5	9.4	8.7
16	1.0	6.4	−0.3	2.1	2.6	1.1	2.3	4.5	9.4	0.2
18	1.7	−2.9	−3.8	−3.2	2.0	10.7	6.3	4.9	8.9	0.2

^a Torsion angles are listed in Table 1. ^b The J-couplings are discussed primarily in relation to the indicated torsion angles.

structural information was taken either from the original work or extracted from the structures as deposited in the NDB.²

Torsion γ . Torsion γ has a trimodal distribution in crystal structures; the dominant peak for double-helical B and A forms has its maximum near 60°, corresponding to “gauche+” conformation (g+). Rarely occurring distributions, trans (t) near 180° and gauche− (g−) near 300°, are represented by the conformational classes 4, 19, and 18, respectively, Table 1. As can be seen in Tables 4 and 5 and Figure 2, J-couplings vary significantly, up to several Hz, with changing values of the torsion. This fact facilitates unequivocal determination of γ . Some J constants undergo specific sign change when switching between the three main conformational regions, g+, t, g−, as shown in Table 6. The regions g+, t, and g− of γ can therefore be determined only from the sign of the corresponding couplings.

A combination of couplings shown in Table 6 is an example of the set that discriminates between all three regions. The couplings ${}^4J(P, C3'1)$ and ${}^2J(C4'1, 1H5'1)$ can be used for discrimination between g± and t conformers. Positive values of ${}^4J(P, H4'1)$, ${}^2J(C5'1, H4'1)$, and ${}^2J(C4'1, 2H5'1)$ couplings

indicate the g+ conformations in a g± mixture. The ${}^3J(H4'1, 1H5'1)$ coupling of roughly 2 Hz was calculated for all investigated conformations in the g+ orientation of γ , while the conformation 18 in g− region had the value of 10.7 Hz. Such a high value of this coupling may result from a planar arrangement of H4'1–C4'1–C5'1–1H5'1 atoms and O=P–O5'–C5'–C4'–O4' atoms, unique for conformation 18.

Torsion δ , neighboring with γ , modulates behavior of J-couplings determined primarily by γ . Because of the fact that δ correlates with the sugar pucker (see also paragraph on δ), this modulation can be conceptualized in terms of the sugar pucker. As shown in Figure 2A and B, couplings ${}^4J(P, C3'1)$ and ${}^3J(C5'1, H3'1)$ are larger by 1–1.5 Hz in conformations with the C2'-endo sugar pucker compared to conformations with the C3'-endo pucker. A significant effect is also observed for ${}^2J(C5'1, C3'1)$ coupling; its magnitude near 1 Hz in g+ conformations changes sign when going from one to the other sugar pucker mode (Table 5).

Five of eight couplings calculated for torsion γ were compared to the experimental data. For couplings ${}^3J(C5'1, H3'1)$ and ${}^2J(C5'1, H4'1)$ shown in Figure 2B and C, respectively, we

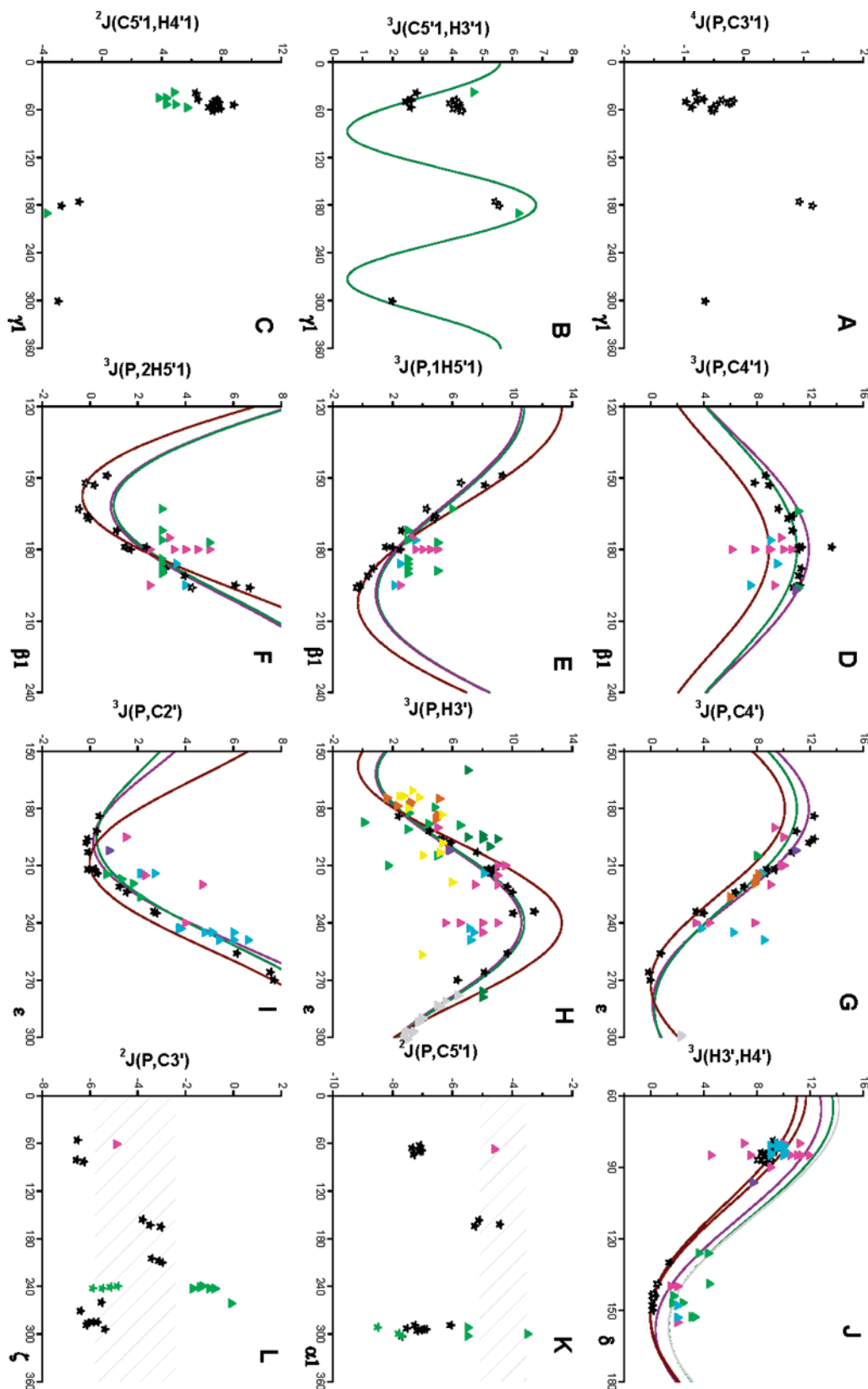


Figure 2. Correlation of J-couplings [Hz] with the backbone torsion angles [degrees]. Stars indicate calculated couplings; in parts A and B, J sugar pucker is indicated by empty stars (C3'-endo) and filled stars (C2'-endo); in parts E and F, α torsion classes indicate empty stars ($\alpha \approx 160^\circ$), filled stars ($\alpha \approx 70^\circ$), and half-filled star ($\alpha \approx 300^\circ$); in part K, green stars indicate calculation in dinucleosides 2, 3, 8, and model 1 in NDB structure 1NEV;²⁷ in part L, green stars indicate calculation in dinucleosides 1, 2, 7, 9, and model 1 in structure with NDB code 1NEV.²⁷ Triangles indicate experimental data from the literature: (B) green;²⁸ (C) green;²⁸ (D) green,²⁹ magenta,³⁰ cyan,³¹ violet;³² (E) green,³³ magenta,³⁰ cyan;³¹ (F) green,³³ magenta,³⁰ cyan;³¹ (G) green,³⁴ magenta,³⁰ cyan,³¹ violet;³² orange,³⁵ gray;³⁶ (H) green,³⁴ magenta,³⁰ cyan,³¹ violet;³² orange,³⁷ gray,³⁸ olive,³³ yellow;³⁹ (I) magenta,³⁰ cyan,³¹ violet;³² orange;³⁵ (J) green,³⁴ magenta,³⁰ cyan,³¹ violet;³² (K) green,³⁴ magenta,³⁰ cyan,³¹ violet;³² gray interval (unknown structure);⁴⁰ (L) green,³⁴ magenta,³² gray interval (unsolved structure).⁴⁰ Karplus curves from the literature: (B);^{6,41} (D) green,⁶ purple,⁸ red;⁴² (E) green,⁶ purple,⁸ red;⁴² (F) green,⁶ purple,⁸ red;⁴² (G) green,⁶ purple,⁸ red;⁴² (H) green,⁶ purple,⁸ red;⁴² (I) green,⁶ purple,⁸ red;⁴² (J) green,⁶ purple,⁸ red;⁴² light gray.⁴³

TABLE 6: Sign of the Coupling Constants for the Three Main Regions of Torsion γ

coupling	conformational region		
	gauche+	trans	gauche−
$^2J(C4'1,1H5'1)^{a)}$	−	+	−
$^2J(C4'1,2H5'1)^{a)}$	+	−	−
$^2J(C5'1,H4'1)^{a)}$	+	−	−
$^4J(P,C3'1)^{b)}$	−	+	−
$^4J(P,H4'1)^{b)}$	+	−	−

^a Ref 44 and this work. ^b This work.

observed a good qualitative agreement between the calculated values and experimental values measured in an RNA hairpin molecule²⁸ and the empirical Karplus curve⁴¹ cited in the review by Marino.⁶ Calculated couplings also correlate well with the experimental couplings $^2J(C4'1,nH5'1)$, $n = 1,2$,²⁸ $^2J(C5'1,-H4'1)$ ²⁸ and $^3J(H4'1,nH5'1)$, $n = 1,2$.⁴⁵

The magnitude of the calculated $^4J(P,H4'1)$ coupling is comparable with the findings by Altona and co-workers⁴⁶ that report a value of ≤ 3 Hz and Reid and co-workers;⁴⁷ in the latter case, the authors did not include this coupling in their analysis “due to an ambiguity in its relationship with γ angle”. The calculated dependence of the $^4J(P,H4'1)$ coupling on γ torsion acquire its negative value of -1 Hz with $g-$ and t regions of γ , while the $g+$ region corresponds to a positive value between 3.5 and 4.6 Hz. The J-couplings of the $g+$ region further cluster with δ torsion, their larger magnitudes in classes 1, 6, and 9 corresponding to the C2'-endo sugar pucker, see the following paragraph.

Torsions δ . Torsion δ is defined by atoms that formally belong both to the backbone and the (deoxy)ribose sugar ring; it thus links conformations of these two nucleic acid constituents. Torsion δ has two main populations, near 80° and 140° , reflecting a strong preference of (deoxy)ribose ring to either of the two pucker modes, C3'-endo and C2'-endo, respectively. The model compound $n-P-n$ contains two nonequivalent torsions δ , labeled δ and $\delta 1$ in Figure 1.

While no ^{31}P coupling can be directly correlated with torsion δ (Table 4), several, especially those that are primarily correlated with ϵ and γ , are modulated by δ . The $^3J(H3',H4')$ coupling shown in Figure 2J correlates with the sugar pucker mode. Its values obtained for the C3'-endo and C2'-endo sugar fall in the range 8.0–9.5 Hz and 0.1–0.3 Hz, respectively. This well-resolved clustering agrees with a number of experiments,^{30,34,47} and experimental values of the $^3J(H3',H4')$ coupling that differ by 5 Hz or more are usually assigned to different sugar puckers.⁴⁴ Both theory and experiment correlate well with the Karplus curves also shown in Figure 2J.^{6,8,42,43} The negative slope of the Karplus curve for δ values between 80° and 150° agrees well with the calculated and experimental values of $^3J(H3',H4')$.

Another J-coupling, $^3J(H2',H3')$, calculated only in the 5'-end (deoxy)ribose of the $n-P-n$ model, is on average only slightly smaller for the C3'-endo pucker than for the C2'-endo one; in addition, the values measured for the C3'-endo pucker possess a large dispersion.³⁰ This coupling therefore does not seem to be suitable for determination of torsion δ .⁴⁴ The $^3J(H2',H3')$ values are larger in DNA than in RNA, likely a consequence of different sugar rings in these molecules.

Torsion β . Experimental distribution of the torsion angle β is unimodal with one wide, approximately symmetric peak with a maximum near 180° ; wide shoulders reach 60° and 300° . The selected backbone conformations sample the lower part of the distribution with $\beta 1$ in the range from 149° to 196° .

The $^3J(P,C4'1)$ and $^3J(P,nH5'1)$, $n = 1,2$, coupling constants shown in Figure 2D–F are indicative of values of torsion angle $\beta 1$. The calculated values are in good qualitative agreement with the experiment.^{30,31} The three couplings measured in oligonucleotides³¹ follow the curvature of the dependencies. In contrast, the agreement is worse for the couplings measured in a RNA hairpin;³⁰ torsion β was assigned to be in the trans orientation in the original work and uncertainty of the couplings was large. Three empirical Karplus curves fit the model calculation.^{6,8,42} The calculated coupling $^3J(P,C4'1)$ grows from 7.8 to 11.6 Hz in the selected RNA classes, except for the conformational class 18, where the value is 13.6 Hz. Such a high value of this coupling may result from a unique conformational class 18, as mentioned under torsion γ .

Dispersion of calculated dependencies for both $^3J(P, nH5'1)$, $n = 1,2$, couplings shown in Figure 2E and F is notable; the J-couplings obtained for the classes with close values of $\beta 1$ and different values of $\alpha 1$ suggest that torsion $\alpha 1$ modulates these couplings. For example, the $^3J(P,1H5'1)$ coupling calculated in three classes with $\beta 1 \approx 150^\circ$ can be grouped by torsion $\alpha 1$: classes 3 and 7 have $\alpha 1$ in the $g-$ region and coupling has values 8.2 and 9.3 Hz, respectively, while class 1 with $\alpha 1$ in the t region has a lower value of coupling, 6.5 Hz. The values of $^3J(P,1H5'1)$ coupling correlated with torsion $\beta 1$ when $\alpha 1$ is at $g-$ thus appear above the Karplus curves,^{29,35} while the opposite holds for conformations with $\alpha 1$ in the trans region. The split of J-coupling by torsion $\alpha 1$ shown in Figure 2E and F is noticeable when values of $\beta 1$ happen to be in a region with sharp slope of the Karplus curve. This described interplay between the neighboring torsion angles β and α demonstrates a need for multidimensional fitting of Karplus curves.

Torsion ϵ . The distribution of torsion angle ϵ in crystal structures is unimodal, starts by a sharp edge near 180° due to the sugar ring closure, and ends below 300° , the selected conformations having values of $\epsilon 1$ between 184° and 270° .

Calculated dependencies for the couplings $^3J(P,C4')$, $^3J(P,-H3')$, and $^3J(P,C2')$ on torsion ϵ shown in Figures 2G, H and L, respectively, correlate well with a number of experiments referenced in Figure 2. Relatively smooth dependencies calculated for the couplings also agree well with the Karplus parametrizations^{6,8,42} and allow distinguishing between gauche− and trans conformations of torsion ϵ , depending on whether the three couplings are smaller or larger than 5 Hz.⁴⁴

The sugar pucker mode can affect the value of $^3J(P,C4')$ coupling; the values linked to C3'-endo sugar, 12.3, 11.9, and 10.7 Hz, were calculated in classes 19, 2, and 1, respectively, and the value of 11.0 Hz was calculated in class 7 with the C2'-endo sugar (left part of Figure 2G). A reverse effect, the decrease of the coupling by 1.4 Hz for the C3'-endo sugar, was calculated for $^3J(P,H3')$ coupling in conformations 9 and 3. This again signals that reliable empirical parametrization of these couplings should consider the sugar pucker. In contrast, torsion ζ affects $^3J(P,C4')$ coupling to a smaller extent; conformations 19, 2, and 1 with different values of ζ , 196° , 284° , and 56° , respectively, have similar values of the coupling.

The $^3J(P,H3')$ coupling in both DNA and RNA is almost the same. The magnitude of four-bond couplings $^4J(P,H2')$ and $^4J(P,H4')$ was calculated as smaller than 1 Hz in all classes.

Torsion α . Backbone torsion angle α has, similarly to γ , three distinct peaks centered near 60° , 180° , and 300° ; a large majority of conformations around 300° corresponds to the major double-helical conformations B and A. Lack of NMR relevant nuclei in the vicinity of the phosphate group reduces the possibility for observation of classical three-bond couplings for torsion

angles α and ζ . Because also other NMR parameters cannot be straightforwardly related to torsions around the phosphodiester link,⁶ α and ζ , any extension of our knowledge about the relationship between their conformational behavior and J-couplings is desirable.

The calculated coupling ${}^2J(\text{P},\text{C}5'1)$, shown in Figure 2K, can discriminate between the $g\pm$ and t conformational regions of α due to its sizable split of approximately 2 Hz. The dispersion of the coupling within each conformational region is, however, large, probably indicating a complicated nature of correlations with the neighboring torsions and, at the same time, complicates its "clear structural interpretation".⁴⁰

As can be seen in Figure 2K, the dispersion of the measured values of the ${}^2J(\text{P},\text{C}5'1)$ coupling in a DNA oligonucleotide³⁴ is also about 2 Hz; the three measured values are 5.5, 5.5, and 3.5 Hz, and their estimated errors are 0.2, 0.3, and 1 Hz, respectively. The calculated values of the J-coupling in $n\text{--P--}n$ are likely to be reliable because the experimental errors and the effect of hydration estimated by our calculations (see Tables 2, 3) are of comparable magnitude. The corresponding structure²⁷ (NDB code 1NEV) deposited later has been, however, solved without considering J-couplings, and all torsions α have been tightly restrained to $g\text{--}$ ($300^\circ \pm 30^\circ$). The α values for which the three experimental J-couplings are available are 292° for residue G2, 302° for C3, and 300° for C8.²⁷ The measured value of ${}^2J(\text{P},\text{C}5'1)$, 3.5 Hz, and the assigned structural model at the residue C8, $\alpha = 300^\circ$, seem to contradict each other, and the J value would better correspond to α in the trans region. Our calculations for these three nucleotides in their experimental geometries (taken from 1NEV²⁷) confirm this apparent discrepancy; the calculated values of ${}^2J(\text{P},\text{C}5'1)$ for nucleotides G2, C3, and C8 are -8.5 , -7.7 , and -7.8 Hz, respectively. Another experiment measured a fragment of *E. coli* 23S ribosomal RNA.⁴⁰ Its structure has not yet been solved; the measured values of ${}^2J(\text{P},\text{C}5'1)$ are between 3.5 and 5.1 Hz, in line with the measurement of the cyclic deoxydinucleotide (pApA),³² 4.6 Hz.

Systematically larger magnitude of the calculated coupling ${}^2J(\text{P},\text{C}5'1)$ compared to its experimental values is not at this moment explained, but it could either reflect the limits of our molecular model of the nucleic acid or it may be a consequence of different scaling of different experiments^{34,40} and calculations. A good agreement between experimental and calculated values of the other couplings leads us to believe that the predicted split of about 2 Hz between the $g\pm$ and trans regions is real; a similar split between gauche and trans regions was also observed for ${}^2J(\text{P},\text{C}3'1)$, see the following paragraph.

Calculated coupling ${}^4J(\text{C}3',\text{C}5'1)$ that could potentially correlate with conformations at torsion α had calculated values smaller than 0.2 Hz for all considered classes.

Torsion ζ . Torsion angle ζ has the most complicated distribution of all backbone torsions. A wide distribution is populated between 50° and 330° , the most populated peak near 300° corresponds to ζ values in the double-helical B and A forms. The selected conformational classes (Table 1) cover the whole interval, with ζ values between 60° and 290° .

The ${}^2J(\text{P},\text{C}3')$ coupling can be correlated with torsion angle ζ ; the dependence is similar to that between the ${}^2J(\text{P},\text{C}5'1)$ coupling and torsion α ; the calculated differences between values of the ${}^2J(\text{P},\text{C}3')$ coupling in the $g\pm$ and trans regions of ζ are even slightly larger, approximately 3 Hz. This larger split calculated for ${}^2J(\text{P},\text{C}3')$ coupling is in agreement with the measurement of a fragment of the 23S ribosomal RNA,⁴⁰ where the lower limit for the coupling was 2.4 Hz and the upper limit

5.8 Hz; the measurement of the cyclic deoxydinucleotide (pApA)³² reports a comparable value of 4.9 Hz (Figure 2L). On the other side, this coupling measured in DNA oligonucleotide³⁴ ranges from 0.8 to 1.7 Hz.

When we calculated the ${}^2J(\text{P},\text{C}3')$ coupling using the experimental geometry of nucleotides,^{27,34} we obtained J values that are virtually identical with our calculation for $n\text{--P--}n$ in the gauche conformation (Figure 2L). This apparent change of J values from experimental to calculated values has the same trend as for the ${}^2J(\text{P},\text{C}5'1)$ coupling and torsion α as described in the preceding paragraph.

Empirical distributions of two consecutive torsion angles ζ and $\alpha 1$ show that their simultaneous occurrence in the trans region is almost impossible in both RNA and DNA. This fact provides an additional independent condition for gauging the internal consistency of experimental values of the J-couplings and conformations at ζ and $\alpha 1$. If the split of about 2–3 Hz between values of the ${}^2J(\text{P},\text{C}3')$ and ${}^2J(\text{P},\text{C}5')$ couplings between $g\pm$ and trans regions is genuine, then either α or ζ assigned to the trans region implies the other torsion in $g\pm$ conformation. In terms of the coupling values, a low magnitude of one coupling implies a large magnitude of the other coupling. This hypothesis is supported by measurement of G739, U744, A749, A750, A752, A753, C758, C759 nucleotides from a fragment of 23S rRNA,⁴⁰ but the measurement of nucleotide G757 contradicts the hypothesis.

Discussion

Conformational space of nucleic acids is multidimensional and very complex. Even for a small fragment of dinucleoside phosphate, a systematic mapping of torsion angles along its backbone is impractical, if not impossible. Complications of a "brute force" approach arise from a high dimensionality of the conformational space but even more from the complexity of the relationships between its conformational parameters, mainly torsion angles. To overcome this obstacle, we limited our calculations of J-couplings to representative conformations known to exist in experimental structures. Such conformations are now well described for double-helical DNA and RNA forms and, with a limited accuracy, also for "irregular" RNA.^{14,15} We used a representative subset of these conformational classes (Table 1) for the NMR calculations with the goal to characterize backbone conformations of nucleic acids with a set of specific J-couplings. Such a "fingerprint" of a few spin–spin couplings could uniquely mark individual types of nucleic acid backbone and should help to solve structures of nucleic acids by NMR methods.

When comparing the experimentally determined and calculated values of J-couplings (structure 2, Supporting Information) one should bear in mind that, in many cases, experimental J-couplings were not used for structure determination and corresponding torsion angles were assigned after the fact. This is the case, for instance, for experiments plotted in Figure 2D, E, F. However, even in cases when the accurate correlation between couplings and torsion angles is not known, the measured J-couplings, their magnitudes, and relative values provide valuable information about conformational regions.

According to the nature of their distributions, the backbone torsion angles can be divided into two groups. The first group, formed by δ , γ , α , has bi- or trimodal distributions that are relatively narrow, "sharp", in each torsional region; e.g., δ has two populated regions, one corresponding to the C3'-endo (deoxy)ribose pucker near 85° , the other corresponding to the C2'-endo pucker near 140° . The second group includes torsion

angles with wide, “soft”, distributions; e.g., torsion ϵ has asymmetric distribution between 180° and 300° , β very wide, almost symmetric distribution centered at 180° . Such a division of torsion angles into groups of “soft” (ϵ , β) and “sharp” (δ , γ , partially α) is useful for interpretation of the calculated NMR J-couplings.

A well-known correlation between J-couplings and the value of a torsion angle was originally derived for the ^3J -couplings in carbohydrates by Karplus.⁴⁸ Despite chemically and structurally very different nature of NAs, Karplus equations have also been parametrized for various spin–spin interactions in these molecules, as shown in Figure 2. A complicated nature of torsion angle populations in NAs and their mutual interplay is noticeable as a splitting of otherwise smooth dependencies for J-couplings (Figure 2A, B, E, and F) and may in some cases limit applicability of the classical one-dimensional Karplus equations. Although J-couplings depend primarily on one torsion, they often show modulation by conformations around other torsion angle(s). The couplings that correlate with a soft torsion can be conceptualized using an usual Karplus-like dependence⁴⁸ that is modulated by the neighboring torsions. However, when such a multidimensional behavior occurs for two sharp torsions, their interplay results in a split into separate clusters of J-values (Figure 2B). Couplings correlated with a “sharp” torsion should therefore be parametrized with explicit consideration of other torsion(s), either in a form of a multidimensional Karplus curve or a curve parametrized for the constant value(s) of the neighboring torsion(s). Analogical multidimensional character of Karplus dependencies has recently been calculated for NMR parameters in L-alanyl-L-alanine⁴⁹ and DNA nucleosides.⁵⁰

The number of distinct well-resolved J-couplings corresponding to individual torsion angles decreases when going from the (deoxy)ribose to the phosphorus atom. The sharp torsions δ and γ can be determined reliably from a number of well-resolved J-couplings based on naturally occurring hydrogen atoms, ^3J -(H, H), and quite often also by other NMR parameters. In some cases, only the sign of these J-couplings is sufficient for determination of the conformational region. The soft torsions, β and ϵ , can be correlated with well-resolved couplings that have a smooth dependence on the torsion, but the resolution of these dependencies can be improved if the effect of the split by the neighboring sharp torsions δ and γ is considered.

The conformational complexity of NAs is most pronounced at the phosphodiester link, described by torsions α and ζ . These torsions cannot be unequivocally determined from correlations with the ^3J couplings, and the couplings $^2\text{J}(\text{P}, \text{C}5'1)$ and $^2\text{J}(\text{P}, \text{C}3')$ can discriminate only between their gauche \pm and trans conformations by a split of 2–3 Hz. This prediction needs to be validated, although it seems to be supported by two independent experimental observations.^{27,40} A more detailed conformational assignment based on these ^2J -couplings is not straightforward because of their large dispersion. This dispersion is likely related to complicated relations between ζ and α and the neighboring torsions, mainly ϵ and β , and can be explained only after we better understand the multidimensional nature of the NA conformational space. It should be mentioned that reliability of conformational determination of torsions α and ζ can, to a limited extent, be enhanced by the fact that one can a priori exclude the trans/trans arrangement at the phosphodiester link and either α or ζ assigned to the trans region thus facilitates assignment of the other torsion.

Assignment of individual torsion angles is an important task for structure determination of short as well as longer sequences of NAs. Tables 4 and 5 can suggest characteristic combinations

of J-couplings that can be used for determination of larger fragments of NAs when a representative set of J-couplings is available. For instance, double-helical forms A and B structurally differ primarily by torsions δ and ϵ (and by torsion χ around the glycosidic bond, not considered in this work); the difference in value of torsion ζ , about 30° , can initially be neglected. These two fundamental forms can therefore be differentiated by distinguishing by couplings $^3\text{J}(\text{P}, \text{X})$, $\text{X} = \text{C}4', \text{H}3'$, and $^3\text{J}(\text{H}2', \text{H}3')$, $^3\text{J}(\text{H}3', \text{H}4')$. A more subtle difference is between two A-RNA conformations, canonical and conformational class 22 (Table 1); they differ in torsions δ , ϵ , and to a lesser extent, in β ; the couplings relevant for their distinction are $^3\text{J}(\text{H}3', \text{H}4')$, $^3\text{J}(\text{P}, \text{X})$, $\text{X} = \text{C}2', \text{C}4'$, and $^3\text{J}(\text{P}, n\text{H}5'1)$, $n = 1, 2$, respectively. More challenging cases include conformational classes with differences at the phosphodiester link as discussed above, e.g., conformations 1, 2, and 7. These non-A RNA conformations occur in structurally important regions: class 1 in purine-rich double-helical stems in non-Watson–Crick base pairs or unpaired, class 2 in short, mostly tetra-, loops of sequence NRNR (N is any, R purine nucleotide), class 7 is the “platform” conformation.⁵¹ These individual structural patterns differ substantially from the canonical as well as minor A-RNA forms and can be recognized only by a set of J-couplings, including those that can determine values of torsions ζ and α , $^2\text{J}(\text{P}, \text{C}3')$ and $^2\text{J}(\text{P}, \text{C}5'1)$. A rare α in the trans region (classes 1, 2) should help determine ζ as gauche \pm , the definitive assignment of ζ to either g+ region (class 1) or g– (class 2) may be problematic, and discrimination between these two classes may rely on a combination of sugar puckers: class 2 has both in C3'-endo, class 1 has a C3'/C2'-endo mixture. Also different values of β in classes 1 and 2 will likely modulate values of the $^2\text{J}(\text{P}, \text{C}3')$ coupling; reliable estimates of ζ however need more extensive calculations and experiments specifically targeted on these more exotic but essential conformations. Class 7 has a similarly low value of β as class 1, but the phosphodiester torsion in the trans region is now ζ and the torsion in limbo is α ; whether α is in + or – gauche can perhaps be determined from modulation of $^2\text{J}(\text{P}, \text{C}5'1)$, but as in the previous case, more work is needed for unequivocal and reliable assignment of all torsions.

On the basis of the previous discussion of the J-couplings between P, C, and H nuclei, we propose the following protocol for assignment of torsion angles in the NA backbone. Torsions δ and γ should be assigned first, perhaps also from experimental evidence other than J-couplings. Their knowledge should then sharpen conformational regions of soft torsions β and ϵ . A combination of all of these four torsions should finally help determine the two most difficult torsions, α and ζ .

Conclusions

Calculated indirect spin–spin coupling constants between ^{31}P , ^{13}C , and ^1H nuclei can be used for qualitative determination of distinct structural patterns of the nucleic acids backbone. The calculation protocol, including the effect of water solvent used in this study, shows a good performance that results in overall good correlation with a number of experiments from the literature. Theoretical predictions of the J-couplings of the NA backbone are generally valid in both RNA and DNA molecules due to the local nature of NMR parameters. Extensive calculations of the J-couplings unveiled in some cases their dependence on several torsion angles; such J-couplings cannot be explained by one-dimensional Karplus equations. The most prominent is the effect of the sugar pucker, measured here by torsion δ . The calculated values of J-couplings and their relationships to the backbone torsion angles can facilitate a more accurate and

reliable structural interpretation of measured J-couplings; a proposed stepwise procedure suggests assignment of the J-couplings from the sugar part to the phosphodiester link, from torsion δ to γ to β to ϵ to α and to ζ . A representative set of J-couplings, the “fingerprint” along the NA backbone, can be interpreted in terms of the overall nucleotide conformation.

Acknowledgment. This work was supported by the Grant Agency of the Czech Republic 203/05/0388 and a grant from NSF DBI 0110076. B.S. and J.S. were supported by a grant LC512 from the Ministry of Education of the Czech Republic.

Supporting Information Available: Geometries for structures 1, 2, 3, 4, 6, 7, 9, 12, 15, 16, 18, 19 (A-RNA), 20 (canonical A-RNA), 22 (A-RNA), A-DNA, BI-DNA, BI. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Sussman, J. L.; Seeman, N. C.; Kim, S.-H.; Berman, H. M. *J. Mol. Biol.* **1972**, *66*, 403.
- (2) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751.
- (3) Noller, H. F. *Science* **2005**, *309*, 1508.
- (4) Zidek, L.; Stefl, R.; Sklenar, V. *Curr. Opin. Struct. Biol.* **2001**, *11*, 275.
- (5) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struct. Biol.* **2004**, *14*, 350.
- (6) Marino, J. P.; Schwalbe, H.; Griesinger, C. *Acc. Chem. Res.* **1999**, *32*, 614.
- (7) Schwalbe, H.; Marino, J. P.; King, G. C.; Wechselberger, R.; Bermel, W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 631.
- (8) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. NMR Spectrosc.* **1998**, *32*, 287.
- (9) Reif, B.; Hennig, M.; Griesinger, C. *Science* **1997**, *276*, 1230.
- (10) Munzarova, M. L.; Sklenar, V. *J. Am. Chem. Soc.* **2002**, *124*, 10666.
- (11) Munzarova, M. L.; Sklenar, V. *J. Am. Chem. Soc.* **2003**, *125*, 3649.
- (12) Kaupp, M.; Malkin, V. G. *J. Comput. Chem.* **1999**, *20*, V.
- (13) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293.
- (14) Murray, L. J.; Arendall, W. B., III; Richardson, D. C.; Richardson, J. S. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13904.
- (15) Schneider, B.; Moravsek, Z.; Berman, H. M. *Nucleic Acids Res.* **2004**, *32*, 1666.
- (16) Schneider, B.; Neidle, S.; Berman, H. M. *Biopolymers* **1997**, *42*, 113.
- (17) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7*; University of California: San Francisco, 2002.
- (18) Sychrovský, V.; Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2000**, *113*, 3530.
- (19) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.
- (20) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR: Basis Principles and Progress*; Springer: Heidelberg, 1990.
- (21) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (22) Buhl, M.; Malkin, V. G. *Calculation of NMR and EPR Parameters*; WILEY-VCH: Weinheim, 2004.
- (23) Ruud, K.; Frediani, L.; Cammi, R.; Mennucci, B. *Int. J. Mol. Sci.* **2003**, *4*, 119.
- (24) Sychrovský, V.; Schneider, B.; Hobza, P.; Zidek, L.; Sklenar, V. *Phys. Chem. Chem. Phys.* **2003**, *5*, 734.
- (25) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **2000**, *104*, 5631.
- (26) Schneider, B.; Patel, K.; Berman, H. M. *Biophys. J.* **1998**, *75*, 2422.
- (27) Barbic, A.; Zimmer, D. P.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 2369.
- (28) Hines, J. V.; Landry, S. M.; Varani, G.; Tinoco, I. *J. Am. Chem. Soc.* **1994**, *116*, 5823.
- (29) Mooren, M. M. W.; Wijmenga, S. S.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Nucleic Acids Res.* **1994**, *22*, 2658.
- (30) Sich, C.; Ohlenschläger, O.; Ramachandran, R.; Groll, M.; Brown, L. R. *Biochemistry* **1997**, *36*, 13989.
- (31) Richter, C.; Reif, B.; Wörner, K.; Quant, S.; Marino, J. P.; Engels, J. W.; Griesinger, C.; Schwalbe, H. *J. Biomol. NMR* **1998**, *12*, 223.
- (32) Blommers, M. J. J.; Haasnoot, C. A. G.; Walters, J.; Vandermarel, G. A.; Vanboom, J. H.; Hilbers, C. W. *Biochemistry* **1988**, *27*, 8361.
- (33) Varani, G.; Cheong, C. J.; Tinoco, I. *Biochemistry* **1991**, *30*, 3280.
- (34) Zimmer, D. P.; Marino, J. P.; Griesinger, C. *Magn. Reson. Chem.* **1996**, *34*, S177.
- (35) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona, C. J. *Biomol. Struct. Dyn.* **1984**, *1*, 1387.
- (36) Lapper, R. D.; Mantsch, H. H.; Smith, I. C. P. *J. Am. Chem. Soc.* **1973**, *95*, 2878.
- (37) Clore, G. M.; Murphy, E. C.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, *134*, 164.
- (38) Sklenar, V.; Bax, A. *J. Am. Chem. Soc.* **1987**, *109*, 7525.
- (39) Wu, Z. R.; Tjandra, N.; Bax, A. *J. Biomol. NMR* **2001**, *19*, 367.
- (40) O'Neil-Cabello, E.; Wu, Z. R.; Bryce, D. L.; Nikonowicz, E. P.; Bax, A. *J. Biomol. NMR* **2004**, *30*, 61.
- (41) Schwarcz, J. A.; Perlin, A. S. *Can. J. Chem.* **1972**, *50*, 3667.
- (42) Yokoyama, S.; Inagaki, F.; Miyazawa, T. *Biochemistry* **1981**, *20*, 2981.
- (43) Haasnoot, C. A. G.; Deleeuw, F.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
- (44) Varani, G.; Aboulela, F.; Allain, F. H. T. *Prog. NMR Spectrosc.* **1996**, *29*, 51.
- (45) Marino, J. P.; Schwalbe, H.; Glaser, S. J.; Griesinger, C. *J. Am. Chem. Soc.* **1996**, *118*, 4388.
- (46) Mellema, J. R.; Pieters, J. M. L.; Vandermarel, G. A.; Vanboom, J. H.; Haasnoot, C. A. G.; Altona, C. *Eur. J. Biochem.* **1984**, *143*, 285.
- (47) Kim, S. G.; Lin, L. J.; Reid, B. R. *Biochemistry* **1992**, *31*, 3564.
- (48) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.
- (49) Bour, P.; Budesinsky, M.; Spirko, V.; Kapitan, J.; Sebestik, J.; Sychrovský, V. *J. Am. Chem. Soc.* **2005**, *127*, 17079.
- (50) Sychrovský, V.; Müller, N.; Schneider, B.; Smrečki, V.; Spirko, V.; Sponer, J.; Trantirek, L. *J. Am. Chem. Soc.* **2005**, *127*, 14663.
- (51) Cate, J. H.; Gooding, A. R.; Podell, E.; Zhou, K.; Golden, B. L.; Szwedczak, A. A.; Kundrot, C. E.; Cech, T. R.; Doudna, J. A. *Science* **1996**, *273*, 1696.

Theoretical Study of the Scalar Coupling Constants across the Noncovalent Contacts in RNA Base Pairs: The *cis*- and *trans*-Watson–Crick/Sugar Edge Base Pair Family

Zuzana Vokáčová,[†] Jiří Šponer,[‡] Judit E. Šponer,[‡] and Vladimír Sychrovský^{*,†,‡}

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo Square 2, 166 10 Prague 6, Czech Republic, and Institute of Biophysics, Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Received: April 11, 2007; In Final Form: June 26, 2007

The structure and function of RNA molecules are substantially affected by non-Watson–Crick base pairs actively utilizing the 2'-hydroxyl group of ribose. Here we correlate scalar coupling constants across the noncovalent contacts calculated for the *cis*- and *trans*-WC/SE (Watson–Crick/sugar edge) RNA base pairs with the geometry of base to base and sugar to base hydrogen bond(s). 23 RNA base pairs from the 32 investigated were found in RNA crystal structures, and the calculated scalar couplings are therefore experimentally relevant with regard to the binding patterns occurring in this class of RNA base pairs. The intermolecular scalar couplings $^1J(\text{N,H})$, $^2J(\text{N,N})$, $^2J(\text{C,H})$, and $^3J(\text{C,N})$ were calculated for the N–H...N and N–H...O=C base to base contacts and various noncovalent links between the sugar hydroxyl and RNA base. Also, the intramolecular $^1J(\text{N,H})$ and $^2J(\text{C,H})$ couplings were calculated for the amino or imino group of RNA base and the ribose 2'-hydroxyl group involved in the noncovalent interactions. The calculated scalar couplings have implications for validation of local geometry, show specificity for the amino and imino groups of RNA base involved in the linkage, and can be used for discrimination between the *cis*- and *trans*-WC/SE base pairs. The RNA base pairs within an isosteric subclass of the WC/SE binding patterns can be further sorted according to the scalar couplings calculated across different local noncovalent contacts. The effect of explicit water inserted in the RNA base pairs on the magnitude of the scalar couplings was calculated, and the data for discrimination between the water-inserted and direct RNA base pairs are presented. The calculated NMR data are significant for structural interpretation of the scalar couplings in the noncanonical RNA base pairs.

Introduction

Nucleic acids (NAs) play a central role in biology. While storage and replication of the genetic information occurs via the structurally rather uniform DNA molecules, RNA molecules are involved in an astonishing range of key biological and biochemical functions, including translation and regulation of gene expression. It appears plausible that a long time ago the world of replicating molecules was dominated by medium-sized RNAs,¹ capable of simultaneously storing and replicating genetic information and performing catalysis.¹

The RNA molecules were later joined by proteins that have more versatile enzymatic capabilities and DNA which acquired the role of storing the genetic information. RNA research has accelerated during the past decade with major achievements such as the discovery of RNA interference^{2,3} and advance of atomic resolution X-ray studies of large RNA assemblies such as the ribosome and its subunits.^{4–7}

Versatility of RNA molecules in biochemical processes stems from their structural variability, and functions of RNA macromolecules are determined by interactions that are by nature noncovalent.

Base pairing belongs to the most important forces in NAs, and its principles are quite different in RNA and DNA.

Hydrogen bonding (H-bonding) is known to stabilize the tertiary structure of NAs, and its detection via NMR spectroscopy would substantially improve the quality of predicted solution NMR structures that are typically rather poorly resolved due to the lack of relevant structural constraints. The classical (Watson–Crick, WC) base pairing dominating in B-DNA is in RNA extended by a highly variable set of non-Watson–Crick (non-WC) interactions that are of the utmost importance for RNA folding and function. For example, about 40% of nucleobases in the ribosome are involved in non-WC pairing and form highly specific and conserved tertiary interactions and recurrent non-WC RNA building blocks.^{8–15}

The role of non-WC base pairs in RNA is clearly fundamental for folding and stabilization of the tertiary structure.

RNA base pairs were recently systemized into 12 basic isosteric families.¹¹ In *cis*- and *trans*-WC/SE (Watson–Crick/sugar edge) RNA base pair families the noncovalent linkage involves one of the relevant functional groups of NA base: the keto, amino, or imino group, and the 2'-hydroxyl group of ribose. Direct detection of specific non-WC base pairs would substantially improve the quality of RNA NMR structures. Many types of RNA interactions were recently investigated by quantum chemical and simulation studies.^{16–26}

Solution structures of NAs can be experimentally determined only indirectly, mostly with the methods of molecular spectroscopy. NMR spectroscopy has been shown to be well-suited for providing reliable structural information of NAs.²⁷ Although

* To whom correspondence should be addressed. E-mail: sychrovsky@uochb.cas.cz.

[†] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

[‡] Institute of Biophysics, Academy of Sciences of the Czech Republic.

the scalar spin–spin coupling constant between two atoms is mostly associated with the respective covalent bond, it can be also measured across the H-bond (the *trans*-H-bond couplings). The dependence of the *trans*-H-bond couplings on molecular structure was recently addressed by Grzesiek,²⁸ and a number of studies of NAs^{29–32} and proteins^{33–36} show its applicability for modeling the molecular structure with the help of NMR parameters. The scalar coupling constants can be accurately modeled using advanced theoretical methods.^{37–43} Theoretical studies show that the *trans*-H-bond *J*-couplings can be used as a fingerprint of specific H-bonding⁴⁴ and correlate strongly with the local geometry of the H-bond.⁴⁵ The scalar coupling constant across an H-bond can be thus utilized for reliable validation of intermolecular interactions in NA base pairs and for accurate probing of their structures.

This study is focused on the dependence of the *trans*-H-bond NMR scalar couplings on molecular structure of *cis*- and *trans*-WC/SE RNA base pairs.¹¹ Since the calculations were performed for a variety of representative NA base pairs, the detailed analysis of scalar couplings has also statistical significance for the NMR structural studies of NAs. We examined 32 members of the WC/SE base pair families, 16 for *cis* and *trans* subclasses. Thirteen members of the *cis*-WC/SE class and 10 members of the *trans*-WC/SE class have been already found experimentally by X-ray crystallography. This ranks the WC/SE base pairs among the most important RNA base pair families. We further considered alternative “water mediated” arrangements of selected base pairs, and formation of “amino-acceptor” interactions where the nucleobase amino group nitrogen is pyramidalized and further acts as a weak H-bond acceptor.⁴⁶ We demonstrate how different interactions of RNA bases occurring in the conventional, amino-acceptor, and water mediated base pairs affect the magnitudes of the scalar couplings. Theoretical modeling of these local effects can improve the accuracy of the modeled NMR parameters and thus the reliability of their structural interpretation.

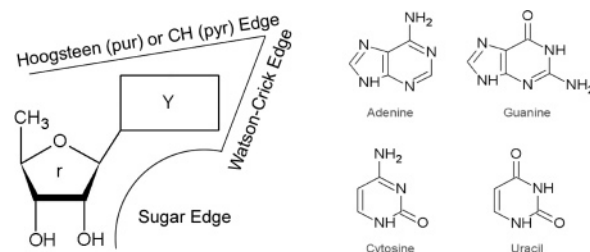
The present study was carried out for the complete ensemble of molecular complexes belonging to the representative class of RNA base pairs. The actual number of RNA base pairs was 42. Therefore, some more general conclusions about the applicability of scalar couplings for distinguishing among individual binding patterns can be made.

Calculation Method

Calculation of the scalar NMR spin–spin coupling constants^{47,48} was performed using the coupled perturbed density functional theory (CP-DFT) method^{49,50} with the B3LYP functional and atomic basis usually called IglolIII,⁵¹ by including all four terms: diamagnetic spin–orbit (DSO), paramagnetic spin–orbit (PSO), Fermi contact (FC), and spin dipolar (SD). The scalar coupling constant ${}^nJ(X,Y)$ was calculated between atoms X and Y, where *n* is the number of covalent bonds separating the linked atoms. For the atoms linked by a noncovalent interaction, the additional index “h” was used, ${}^{nh}J(X,Y)$. The calculations of scalar couplings were performed for the ¹H, ¹³C, and ¹⁵N isotopes.

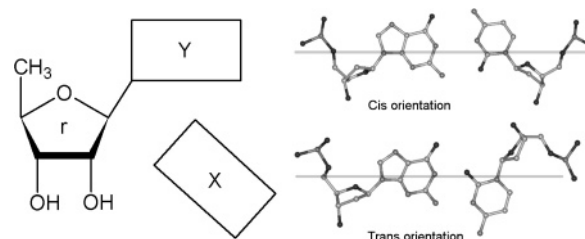
The molecular geometry of the RNA base pairs was optimized previously^{17,18} with the B3LYP functional and 6-31G** basis set. The structural dependence of scalar couplings was calculated in some cases by local variation of the selected geometric parameter, keeping the remaining geometric parameters fixed (the one-dimensional scan). When stated, the polarized continuum solvent (PCM) method⁵² was utilized as a model for inclusion of implicit solvation effects. In all cases the gas phase optimized geometries of RNA base pairs were utilized.

SCHEME 1: Classification of the Interaction Edges for RNA Nucleoside rY by Leontis^{11 a}



^a Y is purine (adenine, guanine) or pyrimidine (cytosine, uracil) base.

SCHEME 2: Schematic Representation of the *cis*- and *trans*-WC/SE RNA Base Pairs^a



^a Interaction between base X and nucleoside rY involves their WC and SE edge, respectively.

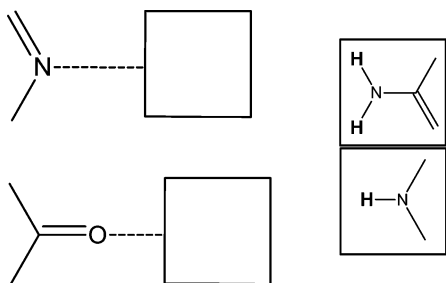
All calculations were done with the Gaussian 03 program package.⁵³

Result and Discussion

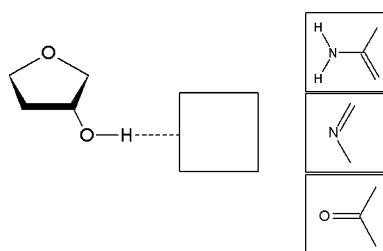
The RNA nucleoside (Scheme 1) possesses three interaction edges: the Watson–Crick edge (WC), the sugar edge (SE), and the Hoogsteen or CH edge defined for purines or pyrimidines, respectively.^{11,54} The sugar edge involves the 2'-hydroxyl group of ribose that is absent in DNA. The RNA nucleoside can interact with another one involving any of the three edges. This interaction can be either *cis* or *trans* (Scheme 2); i.e., the sugars are on the same side or opposite sides of the line linking the interacting nucleotide edges. RNA base pairs were classified using 12 distinct families of well-defined binding patterns by Leontis et al.^{11,54} The isostericity principle suggests that a given base pair can be replaced by another one having a similar (isosteric) shape without a loss of RNA folding and function.^{11,54} The isostericity is approximately reflected by the isostericity index.⁵⁴ The isostericity principle appears to be quite robust, although one should not expect that it is the only constraint determining the covariation patterns in RNAs and significant exceptions certainly exist. In addition, many of the RNA base pairs defined by Leontis et al.^{11,54} can adopt distinct substates while there also exist base pairs that cannot be catalogued based on this classification.

The present study was carried out for the NA bases: adenine (A), cytosine (C), guanine (G), and uracil (U). Typical WC/SE binding patterns (Scheme 2) cluster into the four isosteric subfamilies, I₁, I₂, I₃, and I₄.¹¹ The notation X.rY denotes base X and nucleoside rY interacting via their WC edge and SE edge, respectively. The *cis*- and *trans*-A.rY base pairs (rY is any nucleoside), i.e., those base pairs with adenine interacting via its WC edge, form the subclass having the isostericity I₁ (Table 3 in ref 11). Similarly, the C.rY base pairs are classified by the isostericity indexes I₂ and I₁ for the *cis*- and *trans*-WC/SE base pairs, respectively. The A.rY and C.rY base pairs form isosteric subclasses for their *cis* and *trans* isomers. The behavior of the scalar couplings calculated for the base pairs within isosteric

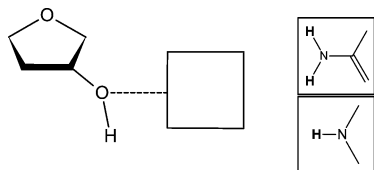
SCHEME 3: Bonding Patterns of the Base to Base Contacts between Ring Nitrogen or Oxygen of Keto Group Involving Amino or Imino Group, the $N\cdots H-N$ Link, and $C=O\cdots H-N$ Link, Respectively



SCHEME 4: Sugar to Base Bonding Patterns between the $O2'-H2'$ Hydroxyl Group and Nitrogen of the Amino Group ($O-H\cdots N(H2)$), Ring Nitrogen ($O-H\cdots N$), and Oxygen of Keto Group ($O-H\cdots O=C$)



SCHEME 5: Sugar to Base Bonding Patterns between the $O2'$ Oxygen of Hydroxyl Group and Hydrogen of Amino or Imino Group ($C-O(H)\cdots H-N$)



classes will be also addressed in the following. We note that the majority of the WC/SE RNA base pairs have been already identified by X-ray crystallography (see the summary in refs 11 and 54).

The calculated scalar couplings were grouped according to the local H-bond patterns. The base to base patterns, i.e., the interaction between ring nitrogen or oxygen of the keto group and the amino or imino group, are shown in Scheme 3. The sugar to base contacts cover a variety of different noncovalent interactions between the $O2'-H2'$ hydroxyl group at the sugar and the relevant site of RNA base; see Schemes 4 and 5.

Since we investigated 42 WC/SE RNA base pairs, all the data are presented in a compact manner as an overall statistics. Dependence of the calculated scalar couplings on geometry is shown in the figures including experimental data extracted from the databases and the literature. All individual calculated scalar couplings can nevertheless be found in the Supporting Information, while base pair geometries are available in the Supporting Information of the preceding papers.^{17,18}

Scalar Coupling Constants across the $N\cdots H-N$ Link. The base to base interactions between the ring nitrogen and hydrogen of the amino or imino group are shown in Scheme 3. The intermolecular ${}^2J(N,N)$ and ${}^1J(N,H)$ scalar couplings across the $N\cdots H-H$ bond occurring in the WC/SE RNA base pairs can be found also in the classical WC/WC base pairs. The intramolecular ${}^1J(H,N)$ couplings calculated for the covalently

linked $N-H$ atoms of the amino or imino group involved in the base to base interactions will be discussed in a separate section together with the ${}^1J(H,N)$ couplings calculated in sugar to base H-bond patterns.

The ${}^2J(N,N)$ couplings calculated in the WC/SE RNA base pairs range from 2.6 to 7.6 Hz (Table 1), in agreement with the interval from 2 to 11 Hz measured in NAs.^{55,56} The dependence of the ${}^2J(N,N)$ couplings on the nitrogen–nitrogen distance in the WC/SE base pairs (Figure 1) is essentially the same as reported for a number of other systems.^{38,43,57,58} The trend calculated here for the ${}^2J(N,N)$ couplings agrees well with the calculations for various $N\cdots H-H$ bonding partners by Del Bene (Figure 1).³⁷ The average value of the calculated ${}^2J(N,N)$ couplings involving an amino group is larger by 1.3 Hz compared to the average for imino contacts (Table 1). The minimal and maximal magnitudes of the ${}^2J(N,N)$ couplings calculated for the amino contacts are larger than those for the imino ones, and both intervals partially overlap. The ${}^2J(N,N)$ couplings measured across the imino contacts in the *cis*-WC/WC DNA base pairs fall into the interval from 5.0 to 8.0 Hz.^{29,31,59,60} The ${}^2J(N,N)$ coupling of 5.5 Hz was measured across the imino contact in the *cis*-Hoogsteen/WC RNA base pair.^{30,61}

Somewhat larger ${}^2J(N,N)$ couplings in WC/WC DNA patterns probably correspond to more favorable geometry as regards the spin–spin coupling pathway. The more collinear are the link atoms at a given interatomic distance, the larger magnitude of the *trans*-H-bond scalar coupling was detected.^{38,45,62,63} Non-WC RNA base pairs are complex H-bonded systems, and not all of their individual H-bonds can be simultaneously fully optimized.^{17,18} Statistics in WC/SE base pairs shows a large declination from the collinear arrangement of the link atoms (Table 1). The average magnitude of the ${}^2J(N,N)$ coupling across the amino and imino contacts was 5.5 and 4.2 Hz, respectively. For comparison, the coupling 5.5 Hz was calculated in the WC/WC canonical GC base pair having the $d(N\cdots H)$ distance of 2.95 Å and the $a(N\cdots H-N)$ angle of 177.5°.

Notably small ${}^2J(N,N)$ couplings close to 2.4 Hz were measured in the mismatched Hoogsteen/Hoogsteen AA DNA base pair across the two amino contacts (Figure 1), although the nitrogen–nitrogen interatomic distance in the structure refined with NMR spectroscopy (1B3P) extracted from the PDB database was 2.844 and 2.779 Å.⁶⁴ The ${}^2J(N,N)$ couplings calculated in the AA base pair when employing raw NMR geometry from the database were 9.1 and 10.8 Hz, and after the geometric optimization, when the nitrogen–nitrogen distance relaxed to 3.057 and 3.026 Å, the couplings dropped to 5.3 and 6.0 Hz, respectively. The impact of local geometry on the magnitude of calculated ${}^2J(N,N)$ couplings is clearly seen, and the remaining deviation of the calculated couplings from experiment can be most probably attributed to the dynamic motion of the mismatched AA pair. Similar damping of the scalar coupling as expected in this case was measured for the ${}^3J(H2,H3)$ coupling in the AU base pairs of RNA hairpin molecule with different flexibilities.⁶⁵ Specifically, the ${}^3J(H2,H3)$ coupling ranged from 1.6 to 1.8 Hz or from 0.1 to 1.0 Hz depending on whether it was measured in regular well-stacked AU base pairs or in more dynamic regions of RNA hairpin, respectively.⁶⁵

The ${}^1J(N,H)$ couplings calculated in the WC/SE base pairs range from 2.1 to 3.1 Hz (Table 1), and only little larger average magnitudes were obtained for the contacts mediated by the amino group compared with the imino ones. The more tight

TABLE 1: Statistics for Calculated Scalar Couplings and Geometric Parameters of the N···H–N Bond Involving Amino or Imino Group of the Base to Base Contacts in WC/SE RNA Base Pairs^a

	$^2J(\text{N},\text{N})$	$^1J(\text{N},\text{H})$	$d(\text{N}\cdots\text{H})$	$\alpha(\text{N}\cdots\text{H}-\text{N})$	$d(\text{N},\text{N})$
amino ^b					
average	5.5	2.8	2.003	169.0	3.006
MDA	0.5	0.2	0.045	8.6	0.028
max	7.6	3.1	2.127	179.0	3.049
min	4.6	2.4	1.860	149.0	2.894
imino ^b					
average	4.2	2.5	2.020	168.7	3.037
MDA	0.9	0.3	0.047	2.3	0.039
max	5.7	2.9	2.117	173.2	3.131
min	2.6	2.1	1.958	163.5	2.987

^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values of the $^2J(\text{N},\text{N})$ and $^1J(\text{N},\text{H})$ intermolecular couplings [Hz], the $d(\text{N}\cdots\text{H})$ hydrogen bond length [Å], the $d(\text{N},\text{N})$ interatomic distance [Å], and bond angle $\alpha(\text{N}\cdots\text{H}-\text{N})$ [deg] were calculated for the patterns shown in Scheme 3. ^b Statistics for 16 amino and 6 imino contacts.

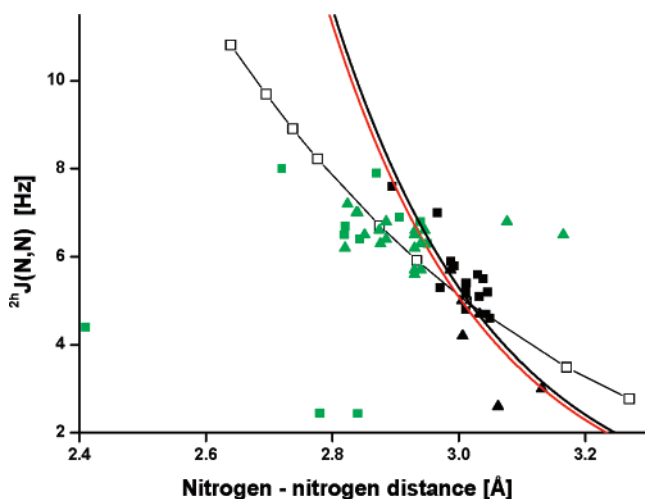


Figure 1. Dependence of $^2J(\text{N},\text{N})$ scalar couplings on the nitrogen–nitrogen distance calculated across the N···H–N link involving amino (■) or imino (▲) group of RNA base in the WC/SE base pairs. Green squares and green triangles stand for experiments involving amino^{32,56,64} or imino^{29,56,59} group, respectively, –□– is the scan calculation for the *cis*-A..A base pair, – is the exponential fit for the WC/SE base pairs, and the red solid line is the exponential fit of the data in various complexes by Del Bene.³⁷

packing of RNA bases results in the increase of the $^1J(\text{N},\text{H})$ coupling as shown for the stepwise change of the nitrogen–nitrogen distance (Figure 2). The relatively large dispersion obtained here for this coupling diminishes its smooth correlation with any geometric parameter. A similar decrease of the $^1J(\text{N},\text{H})$ coupling with a longer H-bond as was calculated in the WC/SE base pairs was nevertheless reported in the literature.³⁸ The $^1J(\text{N},\text{H})$ couplings measured in the WC/WC DNA base pairs range from 1.8 to 4.1 Hz.^{31,40,59,60} The $^1J(\text{N},\text{H})$ couplings measured in the mRNA pseudoknots across the WC/WC contacts range from 1.6 to 2.4 Hz.⁵⁶

The decay of both *trans*-H-bond couplings calculated for the stepwise separation of bases in the *cis*-A..A base pair (Figures 1 and 2) should correspond to some more general trend. Note that the calculated dependencies are ideally smooth since only one geometric parameter was varied. Although dispersion of the calculated $^2J(\text{N},\text{N})$ couplings in the WC/SE base pairs clearly does not exhibit a priori the exponential shape (Figure 1), we used the exponential model for fitting this dependence on the basis of previous theoretical modeling.^{35,37} The exponential dependence $^2J(\text{N},\text{N}) = 795579 \exp[-3.9868(d(\text{N},\text{N}))]$ [Hz] fitted here with the data obtained in the WC/SE base pairs is actually very close to the dependence $^2J(\text{N},\text{N}) = 715763 \exp[-3.9378(d(\text{N},\text{N}))]$ [Hz] by Del Bene³⁷ (Figure 1). Close

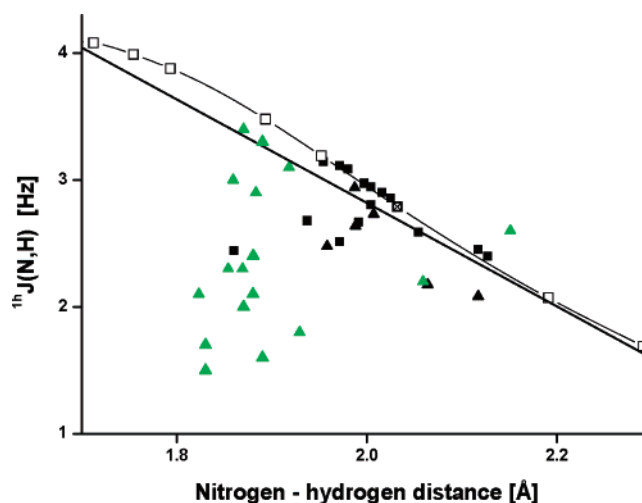


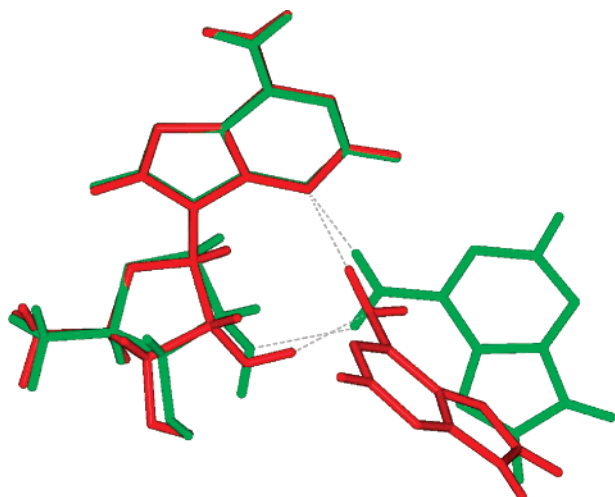
Figure 2. Dependence of $^1J(\text{N},\text{H})$ scalar coupling on the N···H hydrogen bond length calculated across the N···H–N link involving amino (■) or imino (▲) group of RNA base in the WC/SE base pairs. Green triangles stand for experiments involving imino contacts in DNA by Pervushin³¹ and in mRNA by Hennig,⁵⁶ –□– is the scan calculation for the *cis*-A..A base pair, and – is the linear fit for the imino contacts in the WC/SE base pairs.

agreement of both exponential fits indicates that the WC/SE base pair patterns belong to the broad family of the N···H–N intermolecular contacts used for the fitting by Del Bene.³⁷

The scan calculation of the $^1J(\text{N},\text{H})$ coupling (Figure 2) shows its almost linear dependence on the N···H nitrogen–hydrogen distance in the interval from 1.9 to 2.1 Å with the small bend for shorter N···H distances. This agrees with the previous theoretical calculation of the same coupling in NAs correlated with the nitrogen–nitrogen distance of the N–H···N bond (see Figure 5-C in ref 42). The linear dependence $^1J(\text{N},\text{H}) = -4.075(d(\text{N}\cdots\text{H})) + 10.969$ [Hz] (Figure 2) fitted here thus corresponds to the trend calculated previously for the confined interval of the nitrogen–nitrogen distances corresponding to geometries of the WC/SE base pairs.

Both *trans*-H-bond couplings $^2J(\text{N},\text{N})$ and $^1J(\text{N},\text{H})$ seem to be less dependent on the specificity of RNA bases involved in the pairing. The leading factor that correlates well with the coupling magnitude is the distance of bases in the WC/SE RNA pairs.

We also calculated the effect of possible amino group pyramidalization, in the alternative amino-acceptor arrangement. Interestingly, even a relatively large structural perturbation affecting the amino group (Scheme 6) affects the magnitude of the $^2J(\text{N},\text{N})$ and $^1J(\text{N},\text{H})$ couplings less than 0.5 Hz.

SCHEME 6: Overlay of the Conventional (Green) and Amino-Acceptor (Red) *trans*-A.rA WC/SE Base Pairs¹⁸


Scalar Couplings across the C=O...H-N Link. Magnitudes of the $^3J(\text{C},\text{N})$ and $^2J(\text{C},\text{H})$ couplings calculated in the WC/SE base pairs range from -1.2 to 0.1 Hz (Table 2), in agreement with the experiments in proteins.^{33–36,41,66,67} Our calculated data fit the linear correlations $^3J = 2.070(^2J) + 0.420$ [Hz] and $^3J = 1.186(^2J) + 0.113$ [Hz] (Figure 4) for the amino and imino contacts, respectively. The respective standard deviations were 0.15 and 0.08 Hz. The larger standard deviation obtained for the amino contacts corresponds to the larger variation of local geometry (Table 2).

Magnitudes of the $^3J(\text{C},\text{N})$ couplings (Table 2) correspond to the range measured in the DNA guanosine quartets, from -0.18 to -0.26 Hz.³² The $^3J(\text{C},\text{N})$ couplings calculated in the WC/SE base pairs show significant correlation with the length of the H-bond and further with the $\alpha(\text{C}=\text{O}\cdots\text{H})$ bond angle (Figure 3); a decrease in absolute magnitude for a longer H-bond is amplified by a less collinear arrangement of the C=O...H link atoms. A similar strong angular dependence of the $^3J(\text{C},\text{N})$ coupling was also calculated in the methylacetamide dimer,³⁹ and the trend calculated for the $^3J(\text{C},\text{N})$ couplings was measured in proteins.^{34,35}

The slope of the $^3J(\text{C},\text{N})/^2J(\text{C},\text{H})$ correlation in WC/SE base pairs differs for amino and imino contacts. The linear fit of the dependence for the imino contacts calculated here basically coincides with the data measured for the imino contacts in proteins (Figure 4).^{33,36} This indicates that the slope of the $^3J(\text{C},\text{N})/^2J(\text{C},\text{H})$ correlation calculated here across the C=O...H-N link is probably rather conservative in both RNA base pairs and proteins. A different slope of the linear correlation for amino and imino contacts (Figure 4) can be hardly used for their distinguishing since (a) the maximal splitting between the dependencies is approximately 0.5 Hz for rather extreme geometries, which may roughly correspond to a typical experimental error for this coupling, (b) deviations of the calculated data points from the fit are relatively large, and (c) the area of experimental data points in proteins (Figure 4) coincides with the crossing point of both dependencies, which may further diminish the amino/imino resolution for typical C=O...H-N patterns.

Scalar Coupling Constants across the Contacts between Sugar and RNA Base. Several interactions may occur between the O2'-H2' hydroxyl group of ribose and atoms of the WC edge of RNA base. From the RNA base side the interactions can involve the nitrogen of the amino group (C-O-H...N(H₂),

the amino-nitrogen contacts), the ring nitrogen (C-O-H...N, the ring-nitrogen contacts), or the oxygen of the keto group (C-O-H...O=C) (Scheme 4). Alternatively, the O2' oxygen can act as an acceptor of a hydrogen atom of an amino or imino group (C-O(H)...H-N) (Scheme 5). Calculated magnitudes of all *trans*-H-bond couplings separated by three and more bonds were smaller than 0.2 Hz (Supporting Information). This likely rules out utilization of the $^4J(\text{C},\text{C})$ coupling and all $^3J(\text{C},\text{N})$ couplings in NMR structural studies of the sugar to base contacts.

The magnitude of the $^1J(\text{H},\text{N})$ coupling calculated across the C-O-H...N link ranges from 1.2 to 3.6 Hz (Table 3), in agreement with the range from 1.7 to 3.5 Hz measured in RNA by Giedroc et al.⁶⁸ The dependence of the $^1J(\text{H},\text{N})$ coupling on the length of H-bond calculated here differs for amino and ring-nitrogen contacts (Figure 5). The decay for the $^1J(\text{H},\text{N})$ couplings involving the amino group (grid point calculations) is linear with the same slope for *cis*- and *trans*-WC/SE base pairs. The smallest $^1J(\text{H},\text{N})$ coupling (1.2 Hz) was calculated between the H3' hydrogen of sugar hydroxyl and the amino group of guanine only in the *cis*-G.rG base pair having exceptionally long $d(\text{H}3'\cdots\text{N}3)$ bond length (Figure 5). The interaction involving the C3'-hydroxyl should be nevertheless considered an artifact of the model since in RNA the C3'-hydroxyl is replaced by the phosphate group. Other investigated sugar to base contacts involve solely the C2'-hydroxyl.

Closer packing of link atoms along the C-O-H...N bond involving ring nitrogen corresponds to relatively larger magnitudes calculated for the $^1J(\text{H},\text{N})$ couplings (Figure 5). The $^1J(\text{H},\text{N})$ couplings involving ring nitrogen show only modest dependence on the length of H-bond as indicated also by the grid point calculations. The maximal magnitude of the $^1J(\text{H},\text{N})$ coupling was calculated for the intramolecular sugar to base contacts between the N3 ring nitrogen of purine base and the O2'-H2' hydroxyl group of one nucleoside. The couplings 3.6 , 3.2 , and 2.8 Hz were calculated in the *trans*-C.rA, *trans*-U.rA, and *cis*-G.rG base pairs, respectively. The rest of the $^1J(\text{H},\text{N})$ couplings calculated across the intermolecular sugar to base contacts involving ring nitrogen range from 2.3 to 3.1 Hz. Even such a confined interval of the $^1J(\text{H},\text{N})$ couplings can be decomposed according to the bonding partners. The $^1J(\text{H},\text{N})$ couplings involving nitrogen N3 of cytosine (2.3 , 2.3 , 2.3 , and 2.7 Hz) and nitrogen N1 of adenine (2.3 , 2.3 , 2.5 , and 2.9 Hz) (X.rY, X.rC, X.rA, X.rG, X = C and A, respectively) show the gradual increase in the coupling magnitude for the contacts involving purine nucleoside. Further, the $^1J(\text{H},\text{N})$ couplings can involve different ring-nitrogen atoms. Rather specific interchange of the ring-nitrogen atoms can occur for the sugar to base contacts due to the *cis*/*trans* flip of one base in any WC/SE base pair. For example, the *cis* or *trans* coordination of adenine in the A.rX, X = U, C base pairs involves either the N1 or N3 ring nitrogen of adenine. This *cis*/*trans* flip of adenine results in marked increase of the $^1J(\text{H},\text{N})$ coupling from 2.3 to 3.1 Hz.

The intramolecular $^2J(\text{C},\text{H})$ couplings calculated for the hydroxyl group are similar for amino and ring-nitrogen contacts (Table 3). Closer contacts between sugar hydroxyl and ring nitrogen correspond on average to somewhat larger magnitude of the $^2J(\text{C},\text{H})$ coupling. The maximal absolute magnitudes of the $^2J(\text{C},\text{H})$ coupling (-2.5 Hz) were calculated for the contacts between the C2'-hydroxyl and the N3 nitrogen of purine base within one nucleoside (i.e., rG or rA) in the *cis*-G.rG, *trans*-C.rA, and *trans*-U.rA base pairs. Note that the geometry of the rG nucleoside in the *cis*-G.rG base pair is extreme: the

TABLE 2: Statistics for Calculated Scalar Coupling Constants across the C=O...H-N Bond Involving Amino or Imino Group and Geometric Parameters of WC/SE Base to Base Contacts^a

	^{3h} J(C,N)	^{2h} J(C,H)	d(O...H)	a(O...H-N)	a(C=O...H)
amino ^b					
average	-0.6	-0.4	1.952	156.9	139.0
MDA	0.3	0.2	0.096	12.4	10.6
max	0.1	-0.1	2.193	177.2	168.2
min	-1.2	-0.8	1.797	123.5	116.0
imino ^b					
average	-0.4	-0.5	1.959	155.2	133.1
MDA	0.3	0.2	0.165	7.6	7.5
max	0.0	-0.2	2.274	172.5	145.2
min	-0.9	-0.8	1.700	146.6	117.2

^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values of the ^{3h}J(C,N) and ^{2h}J(C,H) intermolecular couplings [Hz], the hydrogen bond length d(O...H) [Å], and the bond angles a(O...H-N) and a(C=O...H) [deg] were calculated for the binding pattern shown in Scheme 3. ^b Statistics for 18 amino and 10 imino contacts.

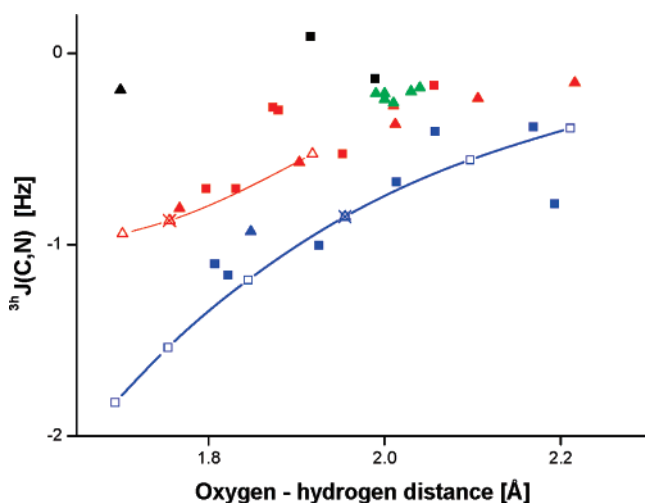


Figure 3. Dependence of ^{3h}J(C,N) scalar coupling on the O...H hydrogen bond length of the C=O...H-N link involving amino (■) or imino (▲) group of RNA base in the WC/SE RNA base pairs; C=O...H angle 115°–122° (black), 122°–141° (red), and 141°–169° (blue). Green triangles stand for experiment for the imino contacts in DNA with C=O...H angle in the range 103°–123° by Dingley,³² blue —□— is the scan calculation in the *cis*-C.rU base pair (C=O...H angle was 150°), and red —△— is the scan calculation in the *trans*-C.rU base pair (C=O...H angle was 138°).

d(O2'–H2') and d(N3...H2') lengths were 1.004 and 1.702 Å. Close contact of sugar hydroxyl with the ring nitrogen accompanied by relative elongation of the O–H bond is thus one of the prerequisites for amplification of the ²J(C,H) coupling.

The ^{2h}J(H,C) couplings calculated across the C–O–H...O=C link between the sugar hydroxyl group and keto group of base range from -0.9 Hz to 0 (Scheme 4, Table 4). Larger magnitudes of the couplings were calculated for a more linear arrangement of the link atoms. The ²J(C,H) couplings calculated for the C–O–H...O=C contacts ranged from -2.3 to -1.5 Hz. A similar trend was calculated for the sugar hydroxyl group involved in C–O–H...N linkage.

All ^{2h}J(C,H) and ^{3h}J(C,N) couplings calculated across the C–O(H)...H–N link involving the amino group (Scheme 5) are smaller than 0.2 Hz, while for the only imino contact that has more a favorable geometry (Table 5) the couplings were -0.4 and -0.2 Hz, respectively.

The ²J(C,H) couplings for sugar hydroxyl involved in the C–O(H)...H–N linkage vary little, from -1.3 to -1.8 Hz. Exceptional ²J(C,H) coupling of -2.5 Hz was calculated in the *trans*-C.rA base pair having sugar pucker C2'-endo. Since the sugar pucker in rest of the nucleosides was C3'-endo, the coupling magnitude seems to be more dependent on the sugar

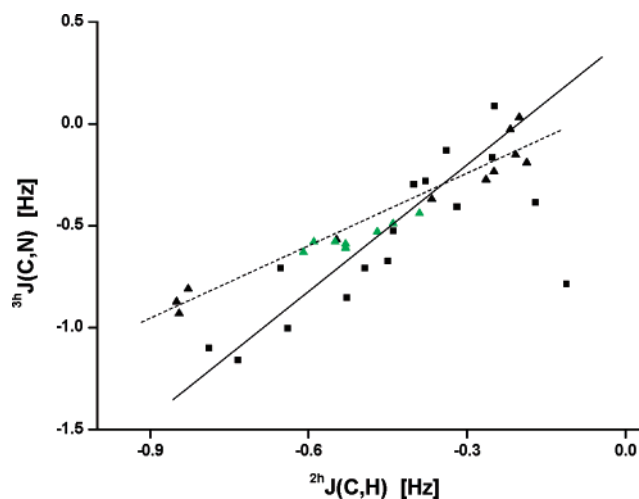


Figure 4. Correlation of ^{3h}J(C,N) and ^{2h}J(C,H) scalar couplings calculated across the C=O...H–N bond involving amino (■) or imino (▲) group of RNA base in the WC/SE RNA base pairs. Green triangles stand for experiments for imino contacts in proteins,³⁶ — is the linear fit for amino contacts, and --- is the linear fit for imino contacts.

pucker than on the geometry of a particular sugar to base contact. A small variation of the ²J(C,H) coupling also corresponds to the weakness of the noncovalent linkage; the C2' hydroxyl is unperturbed by the sugar to base contact and the d(O2'–H2') bond length remains the same (0.973 Å) in all WC/SE base pairs.

^{nh}J(H,H) Couplings across the H-Bond. Scalar spin–spin couplings between the hydrogen atoms taking part in H-bonding offer another constraint for the structural studies in NAs. The ^{3h}J(H2,H3) coupling constants (for numbering of the hydrogen atoms, see Figure 2 in ref 65) measured across the base to base contact in AU WC/WC RNA base pairs range from 0.1 to 1.6 Hz.⁶⁵ The authors reported a strong dependence of the ^{3h}J(H2,H3) coupling on the length of H-bond (FPT approach, only the FC term), and the ^{3h}J(H2,H3) coupling calculated in the AU base pair (~0.1 Hz) was smaller than the experimental couplings measured in the regular part of the RNA hairpin molecule.⁶⁵ Our approach that covers all four spin–spin coupling mechanisms (see Calculation Method) gives a larger magnitude for the ^{3h}J(H2,H3) coupling (0.37 Hz) in the canonical AU WC/WC RNA base pair. The DSO, PSO, FC, and SD coupling contributions were 2.35, -2.14, 0.13, and 0.03 Hz, respectively. The complete calculation including all four terms thus brings the magnitude of the ^{3h}J(H2,H3) coupling in AU base pair closer to the experiment,⁶⁵ and at the same time it points out the necessity of inclusion of all four coupling terms.

TABLE 3: Statistics for Calculated Scalar Coupling Constants and Geometric Parameters of the C–O–H···N Bond between Sugar Hydroxyl Group and Nitrogen of Amino Group or Ring Nitrogen in WC/SE RNA Base Pairs^a

	¹ <i>J</i> (H,N)	² <i>J</i> (C,H)	<i>d</i> (H···N)	<i>a</i> (O–H···N)	<i>a</i> (H···N–C)
amino nitrogen ^b					
average	2.0	−1.9	2.012	157.6	117.3
MDA	0.3	0.1	0.074	8.8	11.1
max	2.5	−1.6	2.292	171.8	138.7
min	1.2	−2.2	1.917	134.1	102.0
ring nitrogen ^b					
average	2.7	−2.2	1.804	164.0	
MDA	0.3	0.3	0.045	6.5	
max	3.6	−1.5	2.031	178.5	
min	2.3	−2.8	1.702	152.4	

^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values were calculated for the ¹*J*(H,N) intercomplex and ²*J*(C,H) intracomplex couplings [Hz]; the *d*(N···H) hydrogen bond length [Å] and the *a*(O–H···N) and *a*(H···N–C) bond angles [deg] were calculated for the sugar to base H-bond pattern shown in Scheme 4. ^b Statistics was calculated for 10 amino-nitrogen and 13 ring-nitrogen sugar to base contacts.

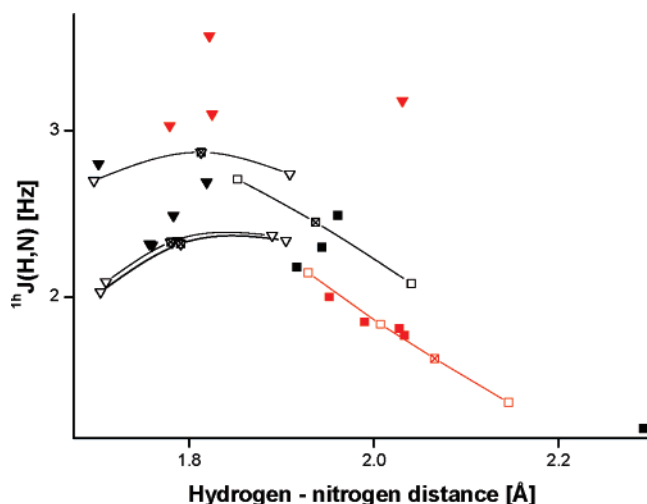


Figure 5. Dependence of intermolecular ¹*J*(H,N) coupling on the hydrogen–nitrogen interatomic distance of the C–O–H···N link between sugar hydroxyl group and amino group (■) or ring-nitrogen atom (▼) in *cis*- (black) and *trans*- (red) WC/SE RNA base pairs. The scan calculations: −▽− for the ring-nitrogen contact in the *cis*-A.rX (X = G > X = U > X = C), −□− for amino contact in the *cis*-G.rA, and red −□− for amino contact in the *trans*-C.rC base pair.

(The H3···N3 bond length in AU base pair used as a geometric descriptor in ref 65 was in our calculation 1.846 Å.).

The astonishing variability of structural patterns in non-canonical RNA base pairs offers many different pathways for the hydrogen–hydrogen scalar couplings across the noncovalent contacts. The ^{*n*}*J*(H,H), *n* = 1, ..., 5, couplings calculated in the WC/SE base pairs are smaller than 0.3 Hz and the majority of the values are close to 0.1 Hz (Supporting Information). The couplings calculated across the base to base contacts, i.e., those between the hydrogen atoms of imino, amino, and C–H groups, are positive and smaller than 0.1 Hz (Supporting Information). The couplings across the sugar to base contacts, i.e., those between the hydrogen atoms of an amino or imino group and sugar hydroxyl, have a little larger magnitude (~0.1 Hz). The maximal coupling (0.27 Hz) was the ²*J*(H,H) one calculated in the *trans*-C.rC amino-acceptor base pair across the contact between the amino group and sugar hydroxyl. Decomposition of both ²*J*(H,H) couplings calculated for the hydrogen atoms of the amino group (0.27 and −0.21 Hz) in the amino-acceptor *trans*-C.rC base pair shows again the complicated nature of this spin–spin interaction: the DSO, PSO, FC, and SD contributions were 1.21, −1.11, 0.09, and 0.07 Hz and 4.15, −4.00, −0.38, and 0.03 Hz, respectively. Interestingly, the same ²*J*(H,H)

couplings calculated in the conventional *trans*-C.rC base pair were only 0.01 and −0.07 Hz.

We can say that reliable theoretical modeling of the ^{*n*}*J*(H,H) couplings requires accurate treatment of all four spin–spin coupling terms due to their strong mutual compensation.

Intramolecular ¹*J*(N,H) Coupling. The ¹*J*(N,H) couplings were calculated for the nitrogen and hydrogen atoms of the amino or imino group. Various noncovalent bonding partners of the amino and imino groups may change local properties of the covalently bound N–H atoms, subsequently affecting magnitudes of the ¹*J*(N,H) coupling. All noncovalent contacts involving the N–H group in the WC/SE base pairs are shown in Schemes 3–5.

The ¹*J*(N,H) couplings calculated in the WC/SE base pairs show a large variation and structurally significant splitting for different bonding patterns of amino and imino groups ranging from −97.5 to −81.5 Hz (Figure 6, Table 6). This sizable variation corresponds to the variation of local geometry from 1.700 to 2.288 Å for the *d*(H···X) length (X = N, O) and from 1.008 to 1.039 Å for the *d*(N–H) length. The variation of the ¹*J*(N,H) couplings calculated for the amino group is larger (16 Hz) compared to the couplings for the imino group (4 Hz). Experimental ¹*J*(N,H) couplings measured for the imino group linked to the ring nitrogen range from −87.5 to −84.7 Hz for RNA and from −88.1 to −85.6 Hz for DNA.⁶⁹

Pyramidalization of the Amino Group. In order to separate the effect of noncovalent bonding on the magnitude of the ¹*J*(N,H) coupling from the intrinsic properties of individual bases, we performed also the calculations in the isolated bases (Table 7). For the amino group of guanine were obtained relatively large absolute magnitudes of the ¹*J*(N,H) couplings compared to the couplings in adenine or cytosine. The amino group of isolated guanine is known to deviate significantly from the planar arrangement.^{70,71} Pyramidalization of the amino group can be quantified by deviation of the sum of the three corresponding valence angles from 360°, and with our particular quantum mechanical method the values are 344°, 357°, and 355° for G, A, and C, respectively. More pyramidal amino group and short *d*(N–H) bond length (Table 7) thus are accompanied by the increase of absolute magnitude of the ¹*J*(N,H) coupling.

Significant deviation of the amino group from planarity found in isolated guanine is suppressed to a large extent in the WC/SE base pairs. The amino groups of adenine (from 344.1° to 359.9°), cytosine (from 345.4° to 360.0°), and guanine (from 351.6° to 357.6°) thus become almost equivalent, and so do the calculated ¹*J*(N,H) couplings in adenine (from −81.5 to −97.5 Hz), cytosine (from −81.8 to −96.7 Hz), and guanine (from −89.5 to −94.1 Hz). Note that the amino groups of bases

TABLE 4: Statistics for Calculated Scalar Coupling Constants and Geometric Parameters of the C–O–H···O=C Bond between Sugar Hydroxyl Group and Oxygen of Keto Group in WC/SE RNA Base Pairs^a

	^{2h} J(H,C)	² J(C,H)	d(H···O)	a(O–H···O)	a(H···N–C)
average ^b	−0.5	−1.9	1.9	164.0	127.5
MDA	0.2	0.2	0.3	11.2	9.5
max	0.0	−1.5	3.4	177.5	146.6
min	−0.9	−2.3	1.7	122.9	107.9

^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values were calculated for the ^{2h}J(H,C) intercomplex and ²J(C,H) intracomplex couplings [Hz]; the d(H···O) hydrogen bond length [Å] and a(O–H···O) and a(H···N–C) bond angles [deg] were calculated for the sugar to base H-bond pattern shown in Scheme 4. ^b Statistics was calculated for 12 sugar to base contacts.

TABLE 5: Statistics for Calculated Scalar Coupling and Geometric Parameters of the C–O(H)···H–N Bond between Sugar Hydroxyl Group and Amino or Imino Group of RNA Base in WC/SE Base Pairs^a

	² J(C,H)	^{2h} J(C,H)	^{3h} J(C,N)	d(O···H)	a(O–H···N)	a(C–O–H)
amino ^b						
average	−1.8	0.0	0.1	2.2	145.6	105.6
MDA	0.2	0.0	0.1	0.1	11.0	0.2
max	−1.3	0.1	0.2	2.3	164.0	106.0
min	−2.5	0.0	−0.1	2.1	131.8	105.3
imino ^b						
cG.rG	−2.8	−0.4	−0.2	1.8	155.5	131.4

^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values were calculated for the ²J(C,H) intracomplex and ^{2h}J(C,H), ^{3h}J(C,H) intercomplex couplings [Hz]; the d(O···H) hydrogen bond length [Å] and the a(O–H···N) and a(C–O–H) bond angles [deg] were calculated for the sugar to base H-bond pattern shown in Scheme 5. ^b Statistics was calculated for 7 amino and 1 imino sugar to base contacts.

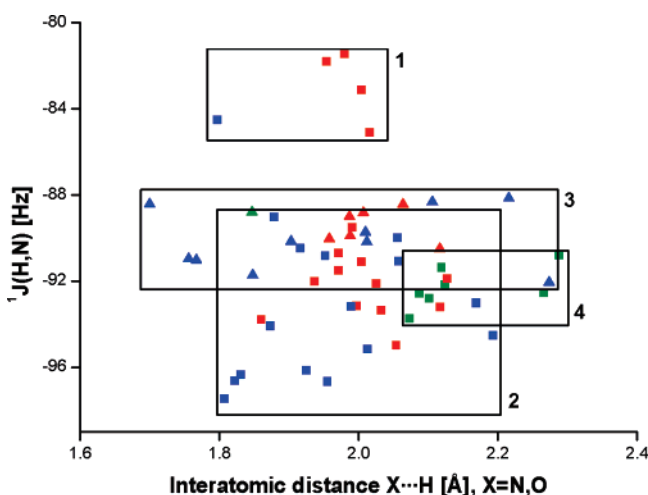


Figure 6. Dependence of ¹J(H,N) coupling on the length of H-bond calculated in the WC/SE RNA base pairs for the C–O(H)···H–N sugar to base contacts (green), and the N···H–N (red) and C=O···H–N (blue) base to base contacts involving amino (■) or imino (▲) group. The regions confine the data points with the same bonding pattern: amino acceptor (1), amino (2), imino (3), and sugar to base (4).

are usually planarized in NAs but there are occasional interactions that actively utilize and even promote amino group pyramidalization—see ref 21 and references therein.

Effect of Hydration. All the above calculations were carried out assuming gas phase conditions since the noncovalent contacts occur inside the base pairs. In order to compare the effect of hydration with the effect of specific contact(s) involving the N–H group, we executed further test calculations by including the solvent effects using the polarized continuum solvent method (see Calculation Method). Polarization of isolated bases by water solvent results on average in the decrease of ¹J(N,H) coupling, at most by 3.6 Hz (Table 7). A similar decrease or increase was calculated when the d(N–H) bond length of the amino or imino group in isolated base shortens or lengthens by 0.02 Å, respectively (data not shown). Variation of the bond length by 0.02 Å corresponds roughly to its change upon differing bonding in WC/SE base pairs (Table 6). The

calculated trends thus indicate that depending on the bonding partner (the geometric effect plus the specific polarization of the N–H group), both effects may amplify or compensate the shift of the ¹J(N,H) coupling due to specificity of the environment. For the *cis*- and *trans*-A.rX and C.rX (X = A, C, G, U) WC/SE base pairs embedded in the implicit solvent, the ¹J(N,H) couplings decrease from 0.1 to 1.4 Hz and from 1.2 to 2.8 Hz for the A.rX and C.rX base pairs, respectively. The larger shift calculated for cytosine most probably corresponds to stronger polarization by solvent in the case of the pyrimidine base. Note that when both N–H bonds of the amino group were completely saturated by two explicit H-bonds, additional implicit hydration of the WC/SE base pair changed the magnitude of the ¹J(N,H) coupling less than 1.2 Hz (*trans*-A.rX (X = A, G), *trans*-C.rX (X = A, C, G, U) base pairs). The calculated ¹J(N,H) couplings are thus predominantly modulated by local geometry in the WC/SE base pairs mostly affected by specific interaction of bonding partners. The effect of solvent most probably leads to the decrease of calculated scalar couplings and should be less pronounced for the base pairs with well explicitly saturated bonding contacts.

Conventional versus Amino-Acceptor Type of Linkage. The amino ¹J(N,H) couplings calculated for the conventional base to base (region 2, Figure 6) and sugar to base (region 4, Figure 6) WC/SE base pairs are well separated from the couplings obtained for those base pairs where the role of donor and acceptor is interchanged and the amino-acceptor interaction is thus formed (region 1, Figure 6).

The topology of the amino-acceptor motif (Scheme 6) assumes the simultaneous occurrence of the base to base linkage involving one N–H bond of the amino group (N···H–N or O···H–N) and the interaction between hydroxyl hydrogen and nitrogen of the amino group. Bases in the conventional WC/SE base pairs are roughly coplanar, while the bases in the amino-acceptor WC/SE base pairs are not, to maximize the interaction between sugar and base. When the amino group undergoes the structural change from the conventional WC/SE base pair to the amino-acceptor one, the ¹J(N,H) coupling calculated for the N–H bond involved in the base to base contact increases: *trans*-A.rA (−93.1 → −81.5 Hz), *trans*-A.rG (−92.1 → −85.1 Hz),

TABLE 6: Statistics for Calculated $^1J(\text{N,H})$ Coupling Constants and Geometric Parameters of the $\text{N-H}\cdots\text{N}$ and $\text{N-H}\cdots\text{O}$ Base to Base Bonding Patterns, and the $\text{N-H}\cdots\text{O}$ Sugar to Base Bonding Pattern in WC/SE Base Pairs^a

group		bonding pattern								
		base to base ^b $\text{N-H}\cdots\text{N}$			base to base ^c $\text{N-H}\cdots\text{O}$			sugar to base ^d $\text{N-H}\cdots\text{O}$		
		$^1J(\text{H,N})$	$d(\text{N}\cdots\text{H})$	$d(\text{H-N})$	$^1J(\text{H,N})$	$d(\text{O}\cdots\text{H})$	$d(\text{H-N})$	$^1J(\text{H,N})$	$d(\text{O}\cdots\text{H})$	$d(\text{H-N})$
amino	average	-89.9	2.003	1.027	-93.1	1.952	1.019	-92.3	2.151	1.011
	MDA	3.6	0.045	0.004	2.8	0.096	0.004	0.7	0.072	0.001
	max	-81.5	2.127	1.035	-84.5	2.193	1.030	-90.8	2.288	1.013
	min	-95.0	1.860	1.020	-97.5	1.797	1.010	-93.7	2.073	1.008
imino	average	-89.4	2.020	1.031	-90.1	1.959	1.026	-88.8	1.847	1.024
	MDA	0.7	0.047	0.004	1.1	0.165	0.006			
	max	-88.4	2.117	1.039	-88.1	2.274	1.036			
	min	-90.5	1.958	1.025	-92.1	1.700	1.019			

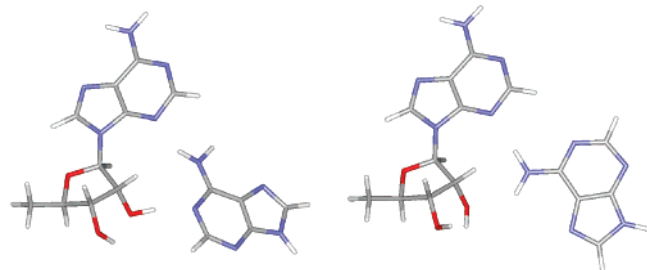
^a Average, mean deviation from average (MDA), and maximal (max) and minimal (min) values were calculated for the $^1J(\text{N,H})$ couplings [Hz]; the $d(\text{N}\cdots\text{H})$, $d(\text{O}\cdots\text{H})$ hydrogen bond lengths [Å] were calculated for the base to base and sugar to base H-bond patterns shown in Schemes 3 and 5. ^b The statistics was calculated for 16 amino and 7 imino base to base contacts. ^c The statistics was calculated for 18 amino and 10 imino base to base contacts. ^d The statistics was calculated for 7 amino and 1 imino sugar to base contacts.

TABLE 7: $^1J(\text{N,H})$ Coupling Constants [Hz] Calculated in Amino-Nitrogen and Imino-Nitrogen Groups of Isolated RNA Bases

base	group	$R(\text{N-H})$	$^1J(\text{N,H})$	$^1J(\text{N,H})^a$
adenine	amino	1.0074	-91.9	-92.7
	amino	1.0067	-92.3	-93.4
guanine	amino	1.0104	-80.1	-83.7
	amino	1.0107	-81.8	-82.5
	imino	1.0128	-87.6	-90.9
cytosine	amino	1.0064	-88.7	-92.0
	amino	1.0089	-89.4	-88.8
uracil	imino	1.0132	-89.8	-89.9

^a NMR calculation including PCM hydration.

SCHEME 7: *cis*-A.rA (Left) and *trans*-A.rA (Right) WC/SE Base Pairs



trans-C.rA (-91.5 \rightarrow -81.8 Hz), *trans*-C.rC (-94.5 \rightarrow -84.5 Hz), and *trans*-C.rG (-91.7 \rightarrow -83.1 Hz).

Analysis of Local Effects in Different Noncovalent Contacts. The $^1J(\text{N,H})$ couplings calculated for the conventional amino WC/SE contacts span the largest interval, and their structural correlation shown in Figure 6 offers only an ambiguous interpretation due to the accumulation of many effects. In the following we discuss these individual effects in connection with the coupling magnitude. First, the amino groups of different bases are inherently not equivalent as was shown in the example of calculation for isolated bases. Analysis of the $^1J(\text{N,H})$ couplings in WC/SE base pairs therefore should be restricted for certain RNA bases. Second, different local bonding partners result in a specific shift of the $^1J(\text{N,H})$ coupling. For example, a smaller average value of the $^1J(\text{N,H})$ coupling was calculated for the $\text{N-H}\cdots\text{O}$ motif compared to the average for the $\text{N-H}\cdots\text{N}$ one (Table 6). This trend goes with the relative shortening of both the H-bond and $d(\text{N-H})$ in the $\text{N-H}\cdots\text{O}$ motif. Third, the local topology of the amino group in one WC/SE base pair having either *cis* or *trans* orientation (Scheme 7) has specific implications for the coupling magnitude. Extended calculation in this work allows statistical analysis of this effect in the subgroups of WC/SE base pairs involving the amino group

of adenine and of cytosine (Figure 7). The $^1J(\text{N,H})$ couplings calculated in the *trans*-WC/SE base pairs are consistently larger than those calculated in the *cis* ones, although the length of the H-bond shrinks ($\text{N-H}\cdots\text{O}$) or lengthens ($\text{N-H}\cdots\text{N}$) when going from *trans*- to *cis*-WC/SE base pairs. Different bonding patterns (Figure 7) are thus more responsible for a larger variation of the amino $^1J(\text{N,H})$ couplings compared to the effect of local geometric variation.

The *cis*- and *trans*-A.rX (X = A, G, C, U) base pairs are isosteric, having the isostericity index I_1 (the classification by Leontis, Table 3 in ref 11). Similarly, the *cis*-C.rX and *trans*-C.rX base pairs are classified by the isostericity indexes I_1 and

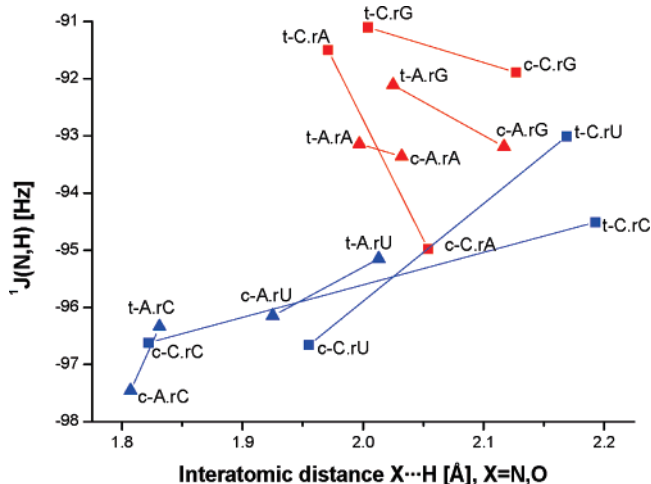
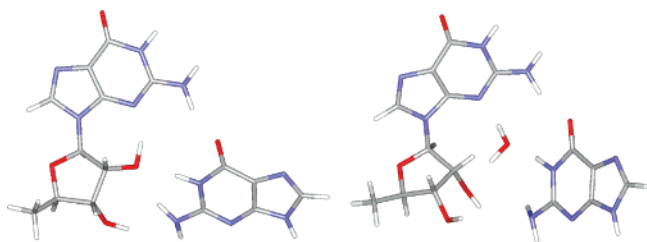
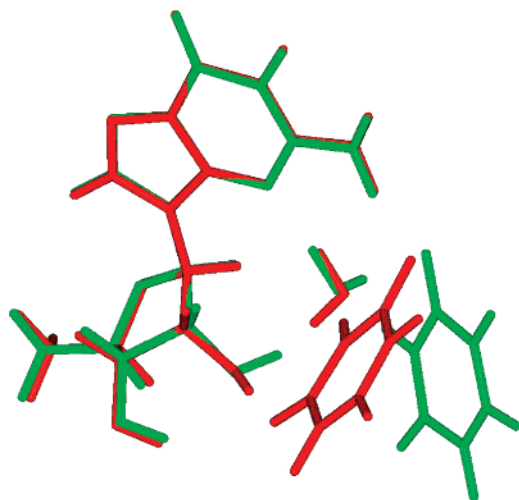


Figure 7. Dependence of $^1J(\text{H,N})$ scalar couplings on the length of H-bond calculated in *cis*- and *trans*-A.rX and -C.rX, X = A, C, G, U, WC/SE RNA base pairs for the amino group of adenine (▲) or cytosine (■) involved in the $\text{N-H}\cdots\text{N}$ (red) and $\text{N-H}\cdots\text{O}$ (blue) linkage. The calculated data points belonging to *cis*- and *trans*-WC/SE base pairs containing the same nucleic acid bases are connected for better clarity.

SCHEME 8: Topology of the Direct and Water Mediated *cis*-G.rG Base Pairs¹⁷**SCHEME 9: Overlay of *trans*-U.rG Base Pairs¹⁸ with Differently Coordinated Water Molecules**

I₂, respectively. The isosteric WC/SE base pairs are considered to be structurally and energetically equivalent. The ¹J(N,H) couplings calculated for the amino group in the isosteric WC/SE base pairs show, however, further clustering for *cis* and *trans* base pairs and according to the local bonding pattern.

The length of the H-bond in sugar to base contacts involving an amino group (region 4, Figure 6) is relatively long, and the amino group is thus relatively unperturbed by the interaction. The calculated ¹J(N,H) couplings are therefore close to the couplings obtained in isolated bases and generally fall into a relatively confined interval (Table 6).

Only modest variation of the ¹J(N,H) couplings was calculated for the imino WC/SE contacts (region 3, Figure 6) with similar average magnitudes for both amino and imino contacts.

NMR Parameters for the Water Mediated RNA Base Pairs. Water molecules can penetrate inside RNA base pairs, having an impact on their structure and stability.^{14,19,20,22,25,26,72–74} In the following we briefly comment on the effect of such an inserted water molecule (or molecules) on the calculated scalar couplings in water mediated RNA base pairs. The *cis*-G.rG pair and two variants of the *trans*-U.rG pair are shown in Schemes 8 and 9. While both hydrated (*cis*-G.rG^w) and water-free (*cis*-G.rG) *cis*-G.rG base pairs can exist, formation of the *trans*-U.rG base pairs requires one water molecule inside the base pair.

The water molecule inside the *cis*-G.rG^w base pair (Scheme 8) hinders the direct contact between the imino group of guanine and sugar hydroxyl. By losing this direct intermolecular linkage, the ¹J(N3,H2') coupling of 2.8 Hz in the *cis*-G.rG pair drops to 0 in the hydrated base pair. At the same time the ²J(C2',H2') coupling increases from −2.8 to −1.2 Hz as a consequence of the change of sugar pucker from the C2'-endo to the C3'-endo. (Note that the C3'-endo is natural for RNA, so the initial C2'-

endo pucker could be an artifact of the initial starting structure taken from a given X-ray base pair or can be related to the gas phase model. On the other side, C2'-endo puckers are not uncommon in noncanonical RNA segments.) The ¹J(N,H) coupling for the amino group linked in both complexes to the keto oxygen increases upon hydration from −94.1 to −89.0 Hz, while the ²J(C,H) and ³J(C,N) couplings remain unchanged, being −0.4 and −0.3 Hz, respectively. The increase due to the hydration calculated for the ¹J(N,H) coupling corresponds to the decrease of planarity of the amino group at guanosine. The sum of the three valence angles calculated for the amino group of guanine was 354.5° and 347.9° in the *cis*-G.rG and *cis*-G.rG^w base pairs, respectively. Other scalar couplings that also respond to the hydration of the *cis*-G.rG base pair are the following: the coupling between the H2' hydrogen of sugar hydroxyl and nitrogen of guanine amino group increases from −0.4 to 1.0 Hz; the coupling between carbon C2' and imino hydrogen drops from −0.4 Hz to 0.

Both hydrated *trans*-U.rG base pairs have similar topology. The scalar couplings between the ring nitrogen N3 of guanosine and the hydrogen or nitrogen of the uracil imino group are smaller than 0.1 Hz, reflecting thus a long distance between the ring nitrogen and imino hydrogen in both *trans*-U.rG base pairs (3.9, 3.4 Å). The magnitudes of the ¹J(H,C) couplings (−0.4, −0.3 Hz) calculated across the N—H···O=C bond are similar to the magnitudes of this coupling calculated in the water-free WC/SE base pairs.

The ¹J(N,H) couplings calculated for the imino group interacting with water molecule in the *cis*-G.rG^w base pair (−87.1 Hz) and in the *trans*-U.rG base pairs (−89.8, −85.9 Hz) indicate the increase of this coupling due to the water insertion. Nonzero ¹J(N,H) couplings were also calculated between the ring nitrogen and directly coordinated hydrogen of the inserted water molecule in the *cis*-G.rG^w base pair (3.1 Hz) and the *trans*-U.rG base pairs (1.8, 3.6 Hz).

Summary of the Calculated Scalar Couplings. The ²J(N,N), ¹J(N,H), ³J(C,N), and ²J(C,H) *trans*-(hydrogen)-bond couplings calculated across the N—H···N and N—H···O=C base to base contacts in the WC/SE base pairs decay with the separation of RNA bases and with declination of the link atoms from collinear arrangement. In this sense the correlation of the *trans*-H-bond scalar couplings in the WC/SE base pairs agrees with theoretical data from the literature^{27,38–44} and available experiments.^{29–36,59–61,64,67,75}

The ²J(N,N) scalar couplings in the WC/SE base pairs have smaller average magnitudes (from 2.6 to 5.7 Hz) than the couplings in the canonical WC/WC G-C DNA base pair (5.5 Hz) due to the large structural variability and less favorable geometry of the WC/SE RNA binding patterns. The nitrogen–nitrogen distance is 2.95 Å and between 2.98 and 3.13 Å in the canonical G-C base pair and in the WC/SE base pairs, respectively (the same optimization method used). The ²J(N,N) couplings involving the amino group calculated in the WC/SE base pairs are on average larger by 1.3 Hz than those involving the imino group, while the ¹J(N,H), ³J(C,N), and ²J(C,H) couplings differ by less than 0.3 Hz.

The ²J(N,N) couplings in the WC/SE base pairs depend on the nitrogen–nitrogen distance exponentially in accord with the benchmark calculation for various bonding partners by Del Bene.³⁷ The scan calculations of the ¹J(N,H) coupling show its linear dependence for the representative geometries in the WC/SE base pairs. Mutual correlation of the ³J(C,N) and ²J(C,H) couplings calculated across the N—H···O=C link is

linear, and the slope differs for the contacts involving the amino and imino groups.

The sugar to base contacts involving the hydroxyl group at sugar can be correlated with the $^1\text{H}J(\text{N},\text{H})$ and $^2\text{H}J(\text{H},\text{C})$ couplings across the $\text{C}-\text{O}-\text{H}\cdots\text{N}$ and $\text{C}-\text{O}-\text{H}\cdots\text{O}=\text{C}$ links, respectively. Other *trans*-hydrogen-bond couplings across three and more bonds between sugar and base in the WC/SE base pairs are smaller than 0.2 Hz.

The $^2J(\text{C},\text{H})$ couplings in the sugar hydroxyl are almost unaffected by sugar to base contacts ranging from -1.3 to -2.8 Hz. Amplification of the $^2J(\text{C},\text{H})$ coupling was calculated only for the $\text{C}2'$ -endo sugar pucker stabilized by the intramolecular contact between sugar hydroxyl and purine base within one nucleoside.

The $^1J(\text{N},\text{H})$ couplings calculated for the amino and imino groups involved in the base to base and sugar to base contacts of the WC/SE base pairs range from -97.5 to -81.5 Hz. A shift by 10 Hz calculated for the $^1J(\text{N},\text{H})$ coupling due to significant nonplanarity of the amino group in isolated guanine is suppressed in the WC/SE base pairs since the amino groups of adenine, cytosine, and guanine involved in the linkages are basically planarized. Simultaneous involvement of one amino group in the base to base ($\text{N}-\text{H}\cdots\text{N}$ or $\text{N}-\text{H}\cdots\text{O}$) and sugar to base ($\text{C}-\text{O}-\text{H}\cdots\text{N}(\text{H}_2)$) linkage (the amino-acceptor interaction) leads to increase of the $^1J(\text{N},\text{H})$ coupling involved in the base to base link by ~ 10 Hz compared to the same couplings calculated in the conventional WC/SE base pairs with the $\text{C}-\text{O}(\text{H})\cdots\text{H}-\text{N}(\text{H})$ sugar to base linkage.

The $^1J(\text{N},\text{H})$ couplings calculated for the *trans*-WC/SE base pairs involving the amino group of adenine and of cytosine are consistently larger than those calculated in the *cis*-WC/SE base pairs. The $^1J(\text{N},\text{H})$ couplings within the *cis*- and *trans*-A.rX ($X = \text{A}, \text{G}, \text{C}, \text{U}$) groups of isosteric WC/SE base pairs thus show also this specific splitting.

Specific insertion of water molecule(s) inside the *cis*-G.rG base pair considered in this work leads to disruption of the intramolecular linkage between the sugar hydroxyl and ring nitrogen of guanosine. The $^1\text{H}J(\text{N},\text{H})$ coupling (2.8 Hz in the direct base pair) drops to 0 upon the water insertion. Insertion of water in the *cis*-G.rG base pair affects sugar pucker ($\Delta^2J(\text{C},\text{H}) \sim 1.6$ Hz) and also planarity of the amino group ($\Delta^1J(\text{N},\text{H}) \sim 5.1$ Hz). When the inserted water molecule hinders intercomplex linkage in the water-inserted *cis*-G.rG and *trans*-U.rG base pairs, the calculated magnitudes of the *trans*-hydrogen-bond couplings are effectively damped.

Conclusion

Correlation of the calculated scalar spin–spin coupling constants in the *cis*- and *trans*-WC/SE RNA base pairs with the molecular structure shows potential for the validation of RNA base pair patterns and structural studies in the noncanonical base pairs with NMR spectroscopy.

Since the existence of majority of the WC/SE base pairs was already confirmed in the solid state and the calculated data are in good agreement with available NMR experiments in nucleic acids and proteins, the presented structural analysis of the scalar couplings should be experimentally relevant. For the *cis*-G.rA, *cis*-U.rU, *trans*-A.rC, *trans*-A.rU, *trans*-C.rU, and *trans*-U.rU WC/SE base pairs that were not yet observed, the calculated data may have implications with regard to their detection.

The base to base contacts in the WC/SE base pairs are detectable with the scalar coupling constants across the $\text{N}-\text{H}\cdots\text{N}$ and $\text{N}-\text{H}\cdots\text{O}=\text{C}$ links. Structural variability of the WC/SE base pairs and frequent deviation of their geometries from ideally

planar patterns contrasting the canonical (WC/WC) base pairs result in effective damping and large dispersion of the calculated scalar couplings.

Validation of the contacts between the sugar hydroxyl and RNA base can provide further critical structural constraints for resolution of the noncanonical RNA base pairs. The $^1\text{H}J(\text{H},\text{N})$ couplings between the hydrogen of sugar hydroxyl and nitrogen of RNA base in the WC/SE base pairs are a little smaller than the couplings calculated across the $\text{N}-\text{H}\cdots\text{N}$ link of the WC/SE base to base contacts. Sugar pucker may affect the magnitude of the $^1\text{H}J(\text{H},\text{N})$ coupling across the sugar to base contacts.

Intramolecular $^1J(\text{N},\text{H})$ couplings calculated for the amino group involved in base to base contacts of the *cis*-WC/SE base pairs are smaller compared to the couplings in the *trans*-WC/SE base pairs. The $^1J(\text{N},\text{H})$ couplings are further specific with regard to the conventional and amino-acceptor geometries of the WC/SE base pair. The calculated data indicate the possibility for the detection of the degree of pyramidalization of the NH_2 group in nucleic acid bases that was already predicted in many systems.^{21,46,76–79}

Insertion of water inside the WC/SE base pair can be of critical importance for the *trans*-hydrogen-bond scalar couplings. Water inside the WC/SE base pairs may saturate essential intermolecular contacts and consequently damps the intercomplex spin–spin couplings to 0. It was shown that also some other intramolecular couplings can respond to the insertion of water.

The base pairs within one isosteric subclass of the WC/SE binding patterns were shown to be further separable according to the scalar couplings that specifically respond to local binding.

Acknowledgment. This work was supported by the Grant Agency of the Czech Republic 203/05/0388 and 203/06/0420, and the Grant Agency of the Czech Academy of Sciences, IAA400550701. This study was further supported by Grants LC06030, MSM0021622413, AVOZZ40550506, and AVOZ-50040507.

Supporting Information Available: Tables of calculated J -couplings and geometric parameters. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA World*, 2nd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1999; Vol. 37.
- Fire, A.; Xu, S. Q.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806.
- Hannon, G. J.; Rossi, J. J. *Nature* **2004**, *431*, 371.
- Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
- Wimberly, B. T.; Brodersen, D. E.; Clemons, W. M., Jr.; Morgan-Warren, R.; Carter, A. P.; Vonnrhein, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327.
- Moore, P. B.; Steitz, T. A. *Annu. Rev. Biochem.* **2003**, *72*, 813.
- Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. *Science* **2005**, *310*, 827.
- Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. *Nucleic Acids Res.* **2005**, *33*, 2395.
- Mokdad, A.; Krasovska, M. V.; Spöner, J.; Leontis, N. B. *Nucleic Acids Res.* **2006**, *34*, 1326.
- Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214.
- Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497.
- Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899.
- Noller, H. F. *Science* **2005**, *309*, 1508.
- Razga, F.; Koca, J.; Spöner, J.; Leontis, N. B. *Biophys. J.* **2005**, *88*, 3466.

- (15) Tamura, M.; Holbrook, S. R. *J. Mol. Biol.* **2002**, *320*, 455.
- (16) Spomer, J. E.; Leszczynski, J.; Sychrovsky, V.; Spomer, J. *J. Phys. Chem. B* **2005**, *109*, 18680.
- (17) Spomer, J. E.; Spackova, N.; Kulhanek, P.; Leszczynski, J.; Spomer, J. *J. Phys. Chem. A* **2005**, *109*, 2292.
- (18) Spomer, J. E.; Spackova, N.; Leszczynski, J.; Spomer, J. *J. Phys. Chem. B* **2005**, *109*, 11399.
- (19) Krasovska, M. V.; Sefcikova, J.; Reblova, K.; Schneider, B.; Walter, N. G.; Spomer, J. *Biophys. J.* **2006**, *91*, 626.
- (20) Spackova, N.; Spomer, J. *Nucleic Acids Res.* **2006**, *34*, 697.
- (21) Spomer, J.; Mokdad, A.; Spomer, J. E.; Spackova, N.; Leszczynski, J.; Leontis, N. B. *J. Mol. Biol.* **2003**, *330*, 967.
- (22) Reblova, K.; Spackova, N.; Stefl, R.; Csaszar, K.; Koca, J.; Leontis, N. B.; Spomer, J. *Biophys. J.* **2003**, *84*, 3564.
- (23) Oliva, R.; Cavallo, L.; Tramontano, A. *Nucleic Acids Res.* **2006**, *34*, 865.
- (24) Das, J.; Mukherjee, S.; Mitra, A.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2006**, *24*, 149.
- (25) Schneider, C.; Brandl, M.; Suhnel, J. *J. Mol. Biol.* **2001**, *305*, 659.
- (26) Brandl, M.; Meyer, M.; Suhnel, J. *J. Phys. Chem. A* **2000**, *104*, 11177.
- (27) Flinders, J.; Dieckmann, T. *Prog. NMR Spectrosc.* **2006**, *48*, 137.
- (28) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. *Prog. NMR Spectrosc.* **2004**, *45*, 275.
- (29) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293.
- (30) Wohnert, J.; Dingley, A. J.; Stoldt, M.; Gorchach, M.; Grzesiek, S.; Brown, L. R. *Nucleic Acids Res.* **1999**, *27*, 3104.
- (31) Pervushin, K.; Ono, A.; Fernandez, C.; Szyperski, T.; Kainosho, M.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14147.
- (32) Dingley, A. J.; Masse, J. E.; Feigon, J.; Grzesiek, S. *J. Biomol. NMR* **2000**, *16*, 279.
- (33) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 1601.
- (34) Cornilescu, G.; Hu, J. S.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 2949.
- (35) Cornilescu, G.; Ramirez, B. E.; Frank, M. K.; Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 6275.
- (36) Cordier, F.; Rogowski, M.; Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1999**, *140*, 510.
- (37) Del Bene, J. E.; Elguero, J. *J. Phys. Chem. A* **2006**, *110*, 7496.
- (38) Alkorta, I.; Elguero, J. *Int. J. Mol. Sci.* **2003**, *4*, 64.
- (39) Scheurer, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **1999**, *121*, 8661.
- (40) Dingley, A. J.; Masse, J. E.; Peterson, R. D.; Barfield, M.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 6019.
- (41) Barfield, M. *J. Am. Chem. Soc.* **2002**, *124*, 4158.
- (42) Barfield, M.; Dingley, A. J.; Feigon, J.; Grzesiek, S. *J. Am. Chem. Soc.* **2001**, *123*, 4014.
- (43) Benedict, H.; Shenderovich, I. G.; Malkina, O. L.; Malkin, V. G.; Denisov, G. S.; Golubev, N. S.; Limbach, H. H. *J. Am. Chem. Soc.* **2000**, *122*, 1979.
- (44) Del Bene, J. E.; Perera, S. A.; Bartlett, R. J. *J. Am. Chem. Soc.* **2000**, *122*, 3560.
- (45) Sychrovsky, V.; Spomer, J.; Trantirek, L.; Schneider, B. *J. Am. Chem. Soc.* **2006**, *128*, 6823.
- (46) Luisi, B. F.; Orozco, M.; Spomer, J.; Luque, F. J.; Shakked, Z. *J. Mol. Biol.* **1998**, *279*, 1123.
- (47) Kaupp, M.; Buhl, M.; Malkin, V. In *Calculation of NMR and EPR Parameters*, Wiley-VCH Verlag: Weinheim, 2004.
- (48) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293.
- (49) Sychrovsky, V.; Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2000**, *113*, 3530.
- (50) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.
- (51) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR—Basis Principles and Progress*; Springer: Heidelberg, 1990.
- (52) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **2000**, *104*, 5631.
- (53) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (54) Leontis, N. B.; Westhof, E. *RNA* **2001**, *7*, 499.
- (55) Grzesiek, S.; Cordier, F.; Dingley, A. J. Scalar couplings across hydrogen bonds. In *Nuclear Magnetic Resonance of Biological Macromolecules, 2001, Part A*; Methods in Enzymology 338; Academic Press: San Diego, 2001; p 111.
- (56) Cornish, P. V.; Giedroc, D. P.; Hennig, M. *J. Biomol. NMR* **2006**, *35*, 209.
- (57) Zidek, L.; Stefl, R.; Sklenar, V. *Curr. Opin. Struct. Biol.* **2001**, *11*, 275.
- (58) Gemmecker, G. *Angew. Chem., Int. Ed.* **2000**, *39*, 1224.
- (59) Pervushin, K. *J. Biomol. NMR* **2001**, *20*, 275.
- (60) Pervushin, K.; Fernandez, C.; Riek, R.; Ono, A.; Kainosho, M.; Wuthrich, K. *J. Biomol. NMR* **2000**, *16*, 39.
- (61) Hennig, M.; Williamson, J. R. *Nucleic Acids Res.* **2000**, *28*, 1585.
- (62) Mishima, M.; Hatanaka, M.; Yokoyama, S.; Ikegami, T.; Walchli, M.; Ito, Y.; Shirakawa, M. *J. Am. Chem. Soc.* **2000**, *122*, 5883.
- (63) Czernek, J.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 11079.
- (64) Majumdar, A.; Kettani, A.; Skripkin, E.; Patel, D. J. *J. Biomol. NMR* **1999**, *15*, 207.
- (65) Luy, B.; Richter, U.; DeJong, E. S.; Sorensen, O. W.; Marino, J. P. *J. Biomol. NMR* **2002**, *24*, 133.
- (66) Li, H.; Yamada, H.; Akasaka, K.; Gronenborn, A. M. *J. Biomol. NMR* **2000**, *18*, 207.
- (67) Alexandrescu, A. T.; Snyder, D. R.; Abildgaard, F. *Protein Sci.* **2001**, *10*, 1856.
- (68) Giedroc, D. P.; Cornish, P. V.; Hennig, M. *J. Am. Chem. Soc.* **2003**, *125*, 4676.
- (69) Manalo, M. N.; Kong, X. M.; LiWang, A. *J. Am. Chem. Soc.* **2005**, *127*, 17974.
- (70) Spomer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.
- (71) Spomer, J.; Hobza, P. *Int. J. Quantum Chem.* **1996**, *57*, 959.
- (72) Holbrook, S. R.; Cheong, C.; Tinoco, I., Jr.; Kim, S.-H. *Nature* **1991**, *353*, 579.
- (73) Cruse, W. B. T.; Saludjian, P.; Biala, E.; Strazewski, P.; Prange, T.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 4160.
- (74) Correll, C. C.; Wool, I. G.; Munishkin, A. *J. Mol. Biol.* **1999**, *292*, 275.
- (75) Majumdar, A. *Magn. Reson. Chem.* **2001**, *39*, S166.
- (76) Tolbert, B. S.; Kennedy, S. D.; Schroeder, S. J.; Krugh, T. R.; Turner, D. H. *Biochemistry* **2007**, *46*, 1511.
- (77) Yildirim, I.; Turner, D. H. *Biochemistry* **2005**, *44*, 13225.
- (78) Vlieghe, D.; Spomer, J.; Van Meervelt, L. *Biochemistry* **1999**, *38*, 16443.
- (79) Spomer, J.; Hobza, P. *J. Am. Chem. Soc.* **1994**, *116*, 709.

Structure and Dynamics of the ApA, ApC, CpA, and CpC RNA Dinucleoside Monophosphates Resolved with NMR Scalar Spin–Spin Couplings

Zuzana Vokáčová,^{*,†} Miloš Buděšínský,^{*,†} Ivan Rosenberg,^{*,†} Bohdan Schneider,[§] Jiří Šponer,[‡] and Vladimír Sychrovský^{*,†,‡}

Institute of Organic Chemistry and Biochemistry, v.v.i., Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic, Institute of Biophysics, v.v.i., Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic, and Institute of Biotechnology AS CR, v.v.i., Vídeňská 1083, 146 20 Prague, Czech Republic

Received: November 5, 2008

The measured NMR scalar coupling constants (J -couplings) in the XpY, (X,Y = adenine (A) or cytosine (C)) RNA dinucleoside monophosphates (DMPs) were assigned to the backbone (α , β , γ , δ , ϵ , ζ) and glycosidic (χ) torsion angles in order to resolve the global structure of the DMP molecules. The experimental J -couplings were correlated with the theoretical J -couplings obtained as the dynamical averages of the Karplus equations relevant to the torsion angles. The dynamical information was captured using the molecular dynamics (MD) calculation method. The individual conformational flexibility of the four DMP molecules was thus consistently probed with the NMR J -couplings. The calculated structure and flexibility of the DMP molecules depend on the sequence considered with respect to the 5' and 3' end of the DMP molecules (5'-XpY-3'). The dynamical characteristics of the two nucleosides are not equivalent even for the ApA and CpC homologues. An enhancement of the sampling in the MD calculations was achieved using five different starting structural motives classified previously for the RNA backbone in the solid phase (Richardson et al. *RNA* 2008, 14, 465–481). The initial structures were selected on the basis of a database search for RNA oligonucleotides. Frequent interconversions between the conformers during the MD calculations were actually observed. The structural interpretation of the NMR spectroscopic data based on the MD simulations combined with the Karplus equations indicates that the dominant conformation of the DMP molecules in solution corresponds to the A-RNA form. For ~52% of the total simulation time (~1000 ns), the $\zeta(g-)-\alpha(g-)-\gamma(g+)$ backbone topology corresponding to the canonical A-RNA form was observed, with roughly equally populated C2'- and C3'-endo sugar pucker interconverting on the nanosecond time scale. However, other noncanonical patterns were also found and thus indicate their relatively high potential to be populated in the dynamical regime. For ~72% of the time portion when the A-RNA of the $\zeta-\alpha-\gamma$ combination occurred, the nucleobases were classified as being mutually stacked. The geometries of the nucleobases classified in this work as stacked were significantly more populated for the DMP molecules with adenosine at the 3' end (ApA and CpA DMPs) than the ApC or CpC RNA molecules with C at the 3' end.

Introduction

Nucleic acids (DNA and RNA) are among the most important classes of biomolecules. DNA is responsible for the storage and coding of genetic information. For quite a long time, the role of RNA was underestimated. However, the astonishing discovery of RNA enzymes (ribozymes) less than 30 years ago¹ signaled a much wider than expected range of RNA's functional capabilities. After the revelation of the RNA catalytic function, major discoveries of new RNA roles and functional capabilities came one after another. These findings revolutionized the biosciences, ultimately identifying the central role of RNA in biology and evolution. We now know that RNA is intimately involved in literally every aspect of cellular life. RNA molecules are integral components of the cellular machinery for protein synthesis and transport, RNA editing and splicing, and chromosome replication and regulation and possess many other

functions. RNA thus functions as an information carrier, catalyst, and regulatory element, most likely reflecting its importance in the earliest stages of evolution (The RNA World).²

Among the recent major advances, we should mention the discovery of the RNA interference³ and the solution of near-atomic resolution X-ray structures of several ribosomal subunits and ribosomes.^{4–11} These structures have provided astonishing insights into the complexity of the architectures of large functional RNAs, the principles of RNA base pairing (strikingly different from DNA), the modularity of functional RNAs, and so forth. X-ray crystallography has been complemented by cryoelectromicroscopy (CryoEM) studies. CryoEM does not achieve atomic resolution but is very versatile and can capture many more conformational substates of this fascinating molecular machine.^{12–14} Notably, the two structural experimental methods above provide basically static averaged pictures of the RNAs and RNA-based molecular machines. However, in order to understand RNA functions and folding fully, dynamical information is needed.¹⁵ It is of primary importance to understand the dynamics and fluctuations of RNAs as molecular machines like ribosomes are stochastic machines working in

* To whom correspondence should be addressed.

[†] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic.

[‡] Institute of Biophysics, Academy of Sciences of the Czech Republic.

[§] Institute of Biotechnology AS CR.

the regime of high viscosity and thermal fluctuations. Biomolecular machines convert the energy of thermal fluctuations into functional movements.¹⁶ Many of the functional capabilities of RNA are based on its chemical properties.

At first glance, these properties may appear rather limited due to the monotony of RNA chemical building blocks, comprising only four nucleotides; nevertheless, RNA typically folds into complex three-dimensional (tertiary) structures. The subtle placement of the functional groups and their structural dynamics determine the RNA's functions. RNA's conformational variability is astonishing and larger than that observed in DNA molecules. RNAs are organized via hierarchical principles. The 2D structure and thermodynamic stability are basically determined by canonical Watson–Crick base pairing of the A-RNA segments.^{17,18} Further, RNA molecules utilize a wide range of noncanonical recurrent structural motifs, building blocks, and tertiary interaction patterns, often characterized by salient sequence, 2D and 3D signatures, which means that they fold into highly specific three-dimensional architectures strongly related to their biological functions.^{19–24}

Vital information on the RNA structure and dynamics can be obtained with NMR spectroscopy,^{25–31} which has become one of the leading techniques in that field. Theoretical MD simulations often complement the experimental insights by a detailed analysis of the stochastic fluctuations and flexibility of RNA molecules on the ~ 1 –1000 ns time scale.^{32–36}

The structural parameters of the nucleic acids resolved with the NMR spectroscopy technique may utilize different NMR parameters.^{37–39} The proton–proton distances can typically be obtained by measuring the nuclear Overhauser enhancement (NOE). The applicability of NOE in nucleic acids is, however, rather limited due to the scarcity of the relevant protons.³⁷ The ¹H, ¹³C, and ³¹P shifts can provide structural information that is unfortunately confined to a close proximity of the probed nucleus. Accurate interpretation of the NMR shifts in terms of long-distance structural restraints may thus become difficult. For example, the ³¹P shifts of the phosphate group vary by only a few ppm despite the large structural variation in the nucleic acid backbone.⁴⁰ One of the long-distance structural restraints in nucleic acids can be obtained with the scalar *J*-couplings measured along a chain of covalently bound atoms⁴¹ or across the noncovalent contacts.^{42,43} The three-bond *J*-couplings (the ³*J*-couplings) correlated with the magnitude of the corresponding torsion angles^{44,45} are thus well suited for structural (conformational) studies of nucleic acids.⁴¹ The specificity of the set of *J*-couplings assigned to individual torsion angles along the RNA backbone with respect to its conformational patterns had been shown also theoretically.⁴⁶ Another long-distance structural restraint can be obtained with the cross-correlated relaxation rates recently utilized in NMR studies of nucleic acids.^{47,48}

This work was motivated by the scarcity of detailed structural information available for short RNA oligonucleotides, which were recognized as increasingly important in the RNA regulatory processes of gene expression briefly mentioned above. The structural characteristics of RNA molecules such as the dependence of flexibility on the sequence (the role of the base-stacking phenomena in the stabilization of short single-strand RNA molecules), and other conditions like various solvent compositions, are essential for understanding the biological function of the molecules. The present joint experimental and theoretical study was therefore particularly focused on (a) the determination of the local structure in terms of torsion angles resolved solely on the basis of measured *J*-couplings (with a consistent set of NMR parameters), (b) a complementary MD study providing

the dynamical interpretation of the measured *J*-couplings in terms of the distribution functions modeled for the torsion angles, (c) the assembly of the local structural information captured for the individual torsions aiming at the global resolution of the RNA conformational patterns (this point can also be classified as a backward validation of the MD calculations method including the force field used), and (d) a detailed analysis of the global RNA topologies resolved with the NMR data for the ApA, ApC, CpA, and CpC RNA dinucleoside monophosphates, addressing particularly the relation between the various topologies of the RNA backbone and base stacking via the sequential permutations of the A (purine) and C (pyrimidine) nucleobases (the A and C nucleobases in the synthesized model compounds were selected on purpose because of their different proposed abilities for base stacking⁴⁹).

Methods Section

The 3',5'-Linked DMP Molecules. These molecules were synthesized in the solid phase on a 30 μ mol scale by a standard phosphoramidite procedure using commercially available amidites and CPG solid supports. After deprotection, the dimers were purified by RP HPLC on a Luna C18 (2) column (5 μ m, 10 \times 250 mm; Phenomenex, U.S.A.) in an aqueous 0.1 M triethylammoniumbicarbonate buffer. The obtained compounds were transformed into sodium salts on Dowex 50 (Na⁺) and finally adjusted by lyophilization from water.

NMR Spectra. The NMR spectra of dinucleoside monophosphates ApA, ApC, CpA, and CpC were measured on a Bruker AVANCE-II instrument (¹H at 600.13 MHz and ¹³C at 150.9 MHz) using a 5 mm TXI cryoprobe in D₂O at a temperature of 280 K (320 K for the CpC DMP molecule). Chemical shifts were obtained from the 1D-¹H and ¹³C NMR spectra, referenced to internal dioxane (1 drop added to D₂O solution) and recalculated to the δ -scale using $\delta_H = 3.75$ and $\delta_C = 69.3$ for dioxane. The homonuclear 2D-H,H-COSY and 2D-H,H-ROESY spectra were used for the structural assignment of hydrogen signals and heteronuclear correlated 2D-H,C-HSQC and 2D-H,C-HMBC spectra for the assignment of the carbon signals. The *J*-coupling constants *J*(H,H) and *J*(H,P) were obtained from 1D resolution-enhanced ¹H spectra, *J*(C,P) from proton-decoupled ¹³C NMR spectra, and vicinal *J*(C,H) between H-1' and the corresponding carbon atom of adenine or cytosine from nondecoupled ¹³C NMR spectra.

Molecular Dynamics. MD simulations were carried out with the AMBER program, version 8,⁵⁰ using the AMBER parm99⁵¹ force field. The RNA molecules were embedded in explicit water (a box of 31 Å \times 31 Å \times 30 Å with roughly 800 TIP3P water molecules). Prior to the MD simulations (50 ns, periodic boundary conditions, a temperature of 300 K, a cutoff at 10 Å), a four-step equilibration was carried out: (1) a solvent relaxation with a rigid geometry of the RNA molecules (500 steps of the steepest descent algorithm followed by 1000 steps of the conjugate gradient minimization), (2) a steepest descent (1000 steps) followed by a conjugate gradient minimization of the RNA backbone (1500 steps), (3) a heating/equilibration simulation at constant volume and a restrained RNA backbone with a temperature increase from 0 to 300 K (20 ps, with the force constant for the backbone torsions being 80 kcal·mol⁻¹·Å⁻², a 2 fs time step, SHAKE algorithm, with the constrained bonds involving hydrogen), and (4) a final equilibration (100 ps, constant pressure, no restraints, SHAKE algorithm). The total simulation time for one DMP molecule was 250 ns (five initial structural patterns with production runs of 50 ns for each starting geometry). For the CpC RNA

molecule, the calculation was performed at a higher temperature of 340 K using the same protocol.

Distribution Functions. The distribution functions for the torsion angles were calculated from their evolution during the MD calculations. The interval of 0–360° for each torsion angle was divided into 36 subintervals of equal size. The number of geometries (snap shots) with the values of the given torsion angle in each subinterval was accumulated along the trajectory (N_i , $i = 1, \dots, 36$). The normalized distributions ($n_i = N_i / (\sum N_i)$) were used for the numerical integration of the Karplus-like equations. It should be mentioned that the calculated values of the averaged J -couplings changed by less than 0.01 Hz when the finer grid (72 subintervals of equal size) was used in the same numerical integration procedure.

Dynamically Averaged J -Couplings. The dynamically averaged J -couplings were calculated using the Karplus equations^{44,45} for nucleic acids.^{41,52–55} The three-bond 3J -couplings were mostly assigned to the corresponding torsion angles. Since several variants of the Karplus equations are available for the individual torsion angles (three for β and ϵ , five for γ and δ) and there is no consensus on which of them is to be preferred, we used their average as reported in the Supporting Information instead. For the two-bond couplings $^2J(C3'1, P)$ and $^2J(P, C5'2)$ assigned to the $\zeta 1$ and $\alpha 2$ torsions, respectively, we used the theoretical model based on our previous calculations⁴⁶ (the model correlates three intervals of 0–120°, 120–240°, and 240–360° for the $\zeta 1$ and $\alpha 2$ torsions with the values of –7.1, –4.9, –7.1 Hz and –6.1, –3.3, –6.1 Hz, respectively). The glycosidic torsion χ was determined with the Karplus equation parametrized theoretically for the $^3J(H1', C)$ couplings by Munzarová and Sklenář⁵⁵ and Wijmenga.⁴¹ The dynamically averaged J -couplings were calculated as numerically integrated Karplus equations with a distribution function of the respective torsion angle ($J = \sum_{i=1}^{36} J(\tau_i) \cdot n_i$, where $J(\tau_i)$ is the middle value of the Karplus equation in the i th subinterval and n_i is the value of the normalized distribution for torsion τ).

Results and Discussion

Selection of the Initial Structures for the MD Calculation.

Short RNA Molecules in the Solid State. RNA molecules containing cytosine and/or adenine (labeled jointly as “C/A” here in this section) were queried in the Nucleic Acid Database (NDB⁵⁶) and compared to the consensus RNA conformers as defined by Richardson et al.⁵⁷ Of the six such C/A dinucleotide structures available in the NDB (ApA: urb003, drb002, ApC: urb001, CpA: drb005, drd004, CpC: udbs38; see Figure 1), two are structures of uncomplexed ribonucleotides (urb003, urb001), but only urb003 has a standard 3'–5' phosphodiester link; the backbone in urb001 contains a 2'–5' link. Three of the dinucleotide structures are C/A complexes with the intercalating drug proflavin (drb002, drb005, drd004), and one contains chemically modified deoxyribose and forms a parallel mini-helix. Also, trinucleotide RNA structures are rare in the NDB; in fact, only one uncomplexed trinucleotide ApApA (urc002) has been deposited, and no C/A RNA tetranucleotide was found.

All of the above-mentioned structures are well-resolved crystal structures providing reliable information about their molecular conformations, but the number of these structures is rather limited and does not allow for a statistical analysis of the observed features. We, therefore, decided to search for short RNA fragments cocrystallized with other RNA and/or proteins in protein/RNA complexes. Approximately 10 such structures were found, including, for instance, the complex between ribonuclease PARN and trinucleotide AAA (pr0068), C₃ and

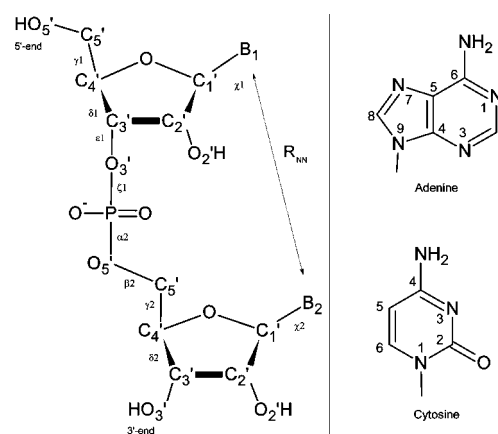


Figure 1. A chemical diagram of the B₁pB₂ DMP molecule (ApA, CpA, ApC, CpC). B₁, B₂ is an adenine or cytosine base of the 5'-B₁pB₂-3' DMP molecule. R_{NN} is the distance between the nitrogen atoms of the bases involved in the glycosidic bond (N1 of C, N9 of A). Indexes “1” and “2” refer to the geometry parameters of the 5'- and 3'-nucleoside; $\gamma 1 = O5'1-C5'1-C4'1-C3'1$, $\delta 1 = C5'1-C4'1-C3'1-O3'1$, $\epsilon 1 = C4'1-C3'1-O3'1-P$, $\zeta 1 = C3'1-O3'1-P-O5'2$, $\alpha 2 = O3'1-P-O5'2-C5'2$, $\beta 2 = P-O5'2-C5'2-C4'2$, $\gamma 2 = O5'2-C5'2-C4'2-C3'2$, $\delta 2 = C5'2-C4'2-C3'2-O3'2$, $\chi = O4'-C1'-N1-C2$ (cytosine), $\chi = O4'-C1'-N9-C4$ (adenosine).

TABLE 1: The Values of the Torsion Angles in Degrees Averaged for the RNA Structural Classes Populated in the Ribosomal RNA⁵⁸ Selected As the Initial Structures in the MD Calculations^a

class ^b	class ^c	$\gamma 1$	$\delta 1$	$\epsilon 1$	$\zeta 1$	$\alpha 2$	$\beta 2$	$\gamma 2$	$\delta 2$	$\chi 1$	$\chi 2$
1a	20	54	81	211	291	296	175	54	84	198	202
1c	19	53	81	194	292	156	194	180	83	203	298
3d	15	53	81	225	211	60	201	49	81	198	196
#a	7	156	149	195	150	292	147	46	84	257	185
6d	9	294	148	237	83	68	175	58	149	56	240

^a These structures were taken as the starting geometries in the MD simulations. ^b New notation used in this work by Richardson et al.⁵⁷ ^c Old notation used in our previous work,⁴⁶ numbering according to Table 3 in ref. ⁵⁸

C₆ oligonucleotides in a complex with a transcription termination factor RHO (pr0008), and several complexes between ribosomal particles and short RNA, such as A₄ mimicking mRNA in a small ribosomal subunit (rr0096) or CCA as an analogue of the 3' end of the deacylated tRNA in a large ribosomal subunit (rr0049). These RNA fragments extended the analysis, but the quality of this data is limited by low resolution while their number (less than 30 C/A steps altogether) would still not allow a rigorous statistical analysis of the conformational behavior of the C/A step in RNA single-stranded stretches of two to four nucleotides.

Despite the relative lack of data, we can draw a qualitative conclusion that the analyzed short single-stranded RNAs containing C and A nucleotides exhibit a high conformational variability. These molecules can acquire conformations from the canonical A-RNA form as the first AA step in urc002, but a variety of other conformers are also populated. Some can be characterized as a deformed A-form, as in the “crankshaft” conformation AII, which is labeled also as consensus conformer 1c (Table 1, Figure 2); other conformations have one sugar pucker C2'-endo rather than the usual C3'-endo, leading to a relatively mildly deformed A-form as in the CA step of drb005 (consensus class 1b). There are also more deformed conformations of the second AA step of urc002 which are close to the consensus class 5z locally resembling a Z-DNA form. Two ApA

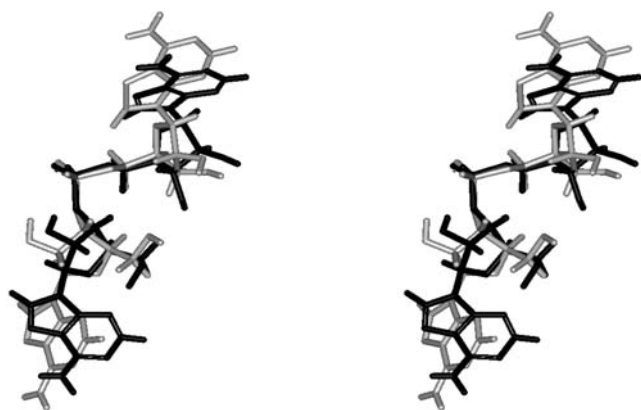


Figure 2. A stereoview overlay of the ApA dinucleoside monophosphate (NDB entry urb003) (black) and the structural class 6d (gray).

dinucleosides, the “naked” urb003 and the proflavine complex drb002, have open and unstacked conformations, while AA in urb003 is close to the class 6d. The conformations of the AA steps found in the complexes with proteins are mostly close to a helical conformation (pr0166, rr0096, pr0068); also, the CCA sequence in the structures of the rRNA subunit (e.g., rr0049) has an almost helical conformation. On the other hand, the CAUC sequence (the NMR structure of the 2i2y fusion protein/RNA complex) is nonstacked with an almost zero rise of bases (“platform” conformers such as #a).

An Ensemble of the Initial Structures for the MD Calculations. The large flexibility of short single-strand RNAs, their open conformations with nonstacked bases, and dissimilarity to the canonical A-RNA form suggest that the conformational space of these molecules should be sampled more comprehensively. The initial structures for the MD calculations should thus include, besides the canonical A-form, at least some of the noncanonical conformers. Therefore, for each sequence, we considered five different starting structures, corresponding to RNA families 1a, 1c, 3d, #a, and 6d (Table 1). The geometries of these conformers were taken from a previous study.⁵⁸ Families 1a and 1c represent the major and minor A-RNA backbone conformations, known also as AI and AII;⁵⁹ #a can be characterized as a platform with a zero rise of bases, and the conformer 3d has unstacked, entirely separated bases; the conformer occurs at the beginnings and ends of the double helices. The use of these five different conformers as the starting structures greatly enhances the sampling of the MD simulations since they cover a variation of all of the important structural features of the backbone angles, base-stacking arrangements, and sugar pucker. We are obviously far from claiming that our five 50 ns simulations for each DMP sequence (20 simulations in total) guarantee convergence of the results. Nevertheless, as shown below, we suggest that our simulation time scale provides enough sampling to achieve a qualitatively meaningful population of the key backbone torsion angles that is sufficient for comparison with the experimental data.

An Analysis of the Molecular Dynamic Calculation. The analysis of trajectories makes it possible to suggest several conclusions. The dynamics calculated for the different torsions occur on specific time scales and possess characteristic torsion angle amplitudes (by amplitude of torsion, we mean here the confined interval of the torsion angle values substantially populated for a given torsion, thus giving rise to a separate peak of the corresponding distribution). For example, the oscillations between the characteristic amplitudes of the torsion angles α , γ , and ζ occur on a scale of tens of ns, whereas in the case of

the δ torsion, the time scale of the flips is around 1 ns. The amplitudes for the χ torsion are broader than those calculated for the α , γ , and ζ backbone torsions, despite their flips among their typical amplitudes occurring on a similar time scale. The specificity of the motions calculated for the structural parameters of the DMP molecules is large. The following discussion is thus focused on the local dynamics of the torsion angles, including their mutual correlations, and on the characterization of the individual distribution functions used for correlating the calculated J -couplings with the NMR experiment. The global structural features like the stacking of bases and their relation to various backbone topologies are discussed separately below.

For the sake of clarity, we have named each simulation in the following way. The simulation with the starting structure N (as indicated in Table 1) of the XpY sequence will be further referred to as NXpY. Thus, for example, Figure 3 shows the full time development and distribution for the #aCpC simulation. All of the other trajectories and distributions can be found in the Supporting Information.

An analysis of the $\alpha 2$, $\gamma 2$, and $\zeta 1$ torsion angles unveils their strong mutual correlation manifested as concerted flips between their typical amplitudes. Each of the three torsion angles has, in principle, a trimodal distribution (g^+ (60°), t (180°), g^- (300°)), but only certain combinations of their amplitudes were populated. The $\alpha 2/g^-, \gamma 2/g^+, \zeta 1/g^-$ triad that pertains to A-RNA (class 1a, i.e. the canonical A-RNA, Table 1) was populated for 52% of the total simulation time, and the second most populated triad $\alpha 2/g^+, \gamma 2/t, \zeta 1/g^+$ was seen for $\sim 19\%$ of the simulation time. Other triads had a fast oscillatory character, and their overall residence time was shorter; they were seen for less than 10% of the total simulation time (Figure 5). The results are in qualitative agreement with the simulations of A-RNA as well as folded RNAs. Specifically, A-RNA simulations with the same force field show the dominance of the canonical g^+/g^- for α/g and a $\sim 10\text{--}15\%$ population of the short-living ($0.1\text{--}20$ ns) g^+/t substates.^{60,61} The crystal structures of the RNA oligonucleotides indicate $\sim 5\%$ population of g^+/t for α/γ . Therefore, the α/γ clearly represent an intrinsic feature of the RNA backbone.^{58,60,61}

The concerted flips of the amplitudes calculated for the $\alpha 2$, $\gamma 2$, and $\zeta 1$ torsions either occur coherently ($\alpha 2/g^-, \gamma 2/g^+, \zeta 1/g^- \rightarrow \alpha 2/g^+, \gamma 2/t, \zeta 1/g^+, \#aApA$) or may include only a subgroup of the torsions ($\alpha 2/g^-, \gamma 2/t, \zeta 1/g^- \rightarrow \alpha 2/g^-, \gamma 2/g^+, \zeta 1/g^- \rightarrow \alpha 2/g^+, \gamma 2/t, \zeta 1/g^+$; 1cApC simulation). Flips in the $\alpha 2$, $\gamma 2$, or $\alpha 2, \zeta 1$ subgroup of torsions were rather frequent, whereas the flip of the $\gamma 2, \zeta 1$ torsions were always accompanied with a flip of the $\alpha 2$ torsion. This may imply that concerted oscillations in the $\zeta 1\text{--}\alpha 2\text{--}\gamma 2$ segment include primarily the $\alpha 2$ torsion (via $\alpha 2 \leftrightarrow \gamma 2$ and/or $\alpha 2 \leftrightarrow \zeta 1$ coupling), while isolated oscillations in the $\zeta 1\text{--}\gamma 2$ subsegment are not allowed.

The $\alpha 2/t$ and $\zeta 1/t$ amplitudes were not populated at all, in agreement with the X-ray data.⁵⁸ Whenever the $\gamma 2/t$ amplitude appeared in a trajectory, both $\zeta 1$ and $\alpha 2$ torsions adopted either g^+ or g^- orientation. The 5'-hydroxyl undergoes a relatively free rotation, and the $\gamma 1$ trajectories show faster oscillations between the $\gamma 1/g^+$ and $\gamma 1/t$ amplitudes than the trajectories of the $\gamma 2$ torsion.

Allow us now to comment on the influence of the initial structures on the sampling. Although the simulations of each DMP were started assuming five different geometries and we generally did see quite many transitions, the influence of the starting structures is visible. The MD calculations with the initial class 1a geometry (canonical A-RNA) retain mostly the initial

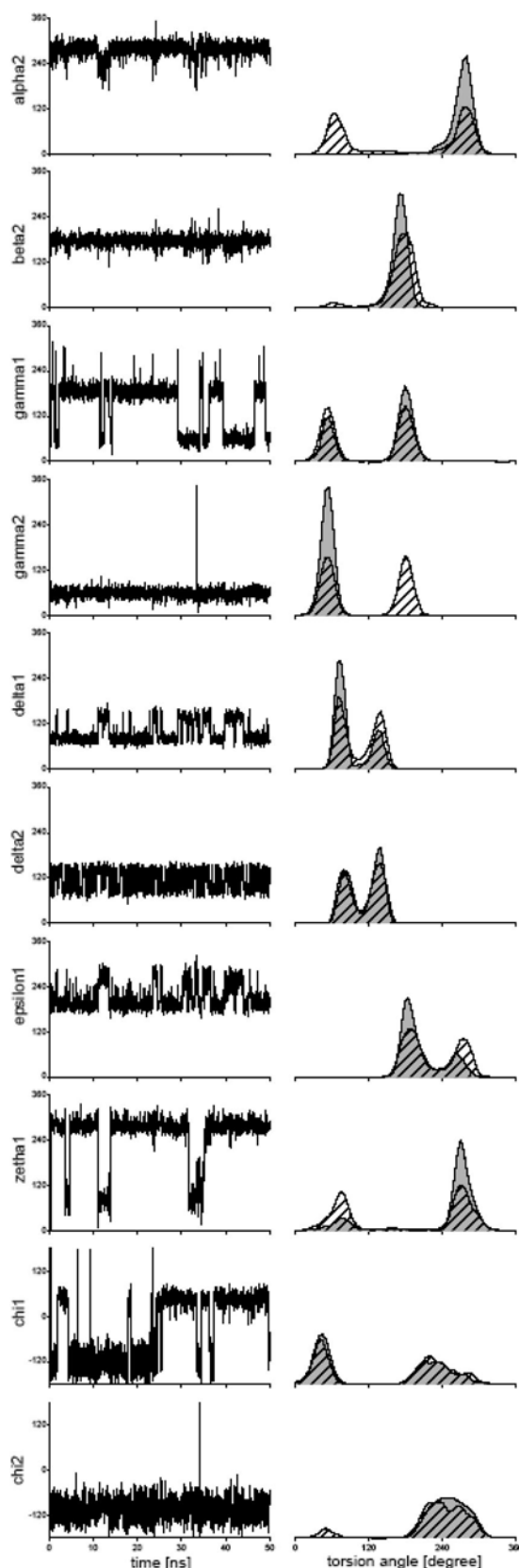


Figure 3. The evolution of the torsion angles over 50 ns and their probability distribution functions for the #aCpC DMP calculated for the temperatures 300 (gray) and 340 K (shaded). The complete data for all of the simulations are in the Supporting Information.

$\alpha 2/g-$ and $\gamma 2/g+$ amplitudes as compared to the stability of those calculated with class 1c, differing only in the $\alpha 2/t$ and $\gamma 2/t$ amplitudes (Table 1, Figure 5). The $\gamma 2/t$ amplitude was

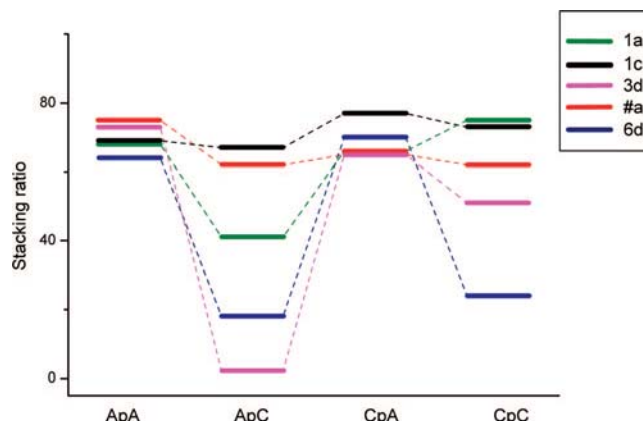


Figure 4. The dependence of the calculated base-stacking ratio (percentage of the geometries with bases in a stack in the course of the simulations) on the sequence of the RNA dinucleoside monophosphates calculated with the five initial structures (inset top right; cf. also Table 2).

populated significantly only in the 1cCpA calculation, a mixture of the $\gamma 2/t, g+$ was found in the 1cApC trajectory, while the 1cApA and 1cCpC trajectories were dominated by the $\gamma 2/g+$ amplitude. This fact may indicate a preference for the canonical A-RNA pattern. The overall tendency for convergence toward the $\alpha 2/g-$, $\gamma 2/g+$ pattern of A-RNA was not observed because in some calculations with different starting structures lasting 50 ns, it was not populated (Figure 5). The initial RNA backbone patterns (Table 1) selected from their complete ensemble⁵⁷ thus represent reliable energy minima with a high individual potential to be populated.

The $\epsilon 1$ torsion angle basically oscillates between the $\sim 180^\circ$ and $\sim 260^\circ$ amplitudes, except for rare and short-lived populations with an $\epsilon 1$ of $\sim 60^\circ$ (6dCpA, 1aApA). The fast oscillations of the $\epsilon 1$ torsion (6d-ApC, CpC) are sometimes damped (1cCpA), and if so, one of the dominant amplitudes is populated more. The coherent oscillations of the $\epsilon 1$ and $\delta 1$ torsions calculated for the 1cXpY, X, Y = A, C, DMP molecules may indicate their possible coupling, and the higher population of $\epsilon 1 \sim 260^\circ$ may thus be correlated with the stabilization of the C2'-endo ribose (1cCpA).

The single $\beta 2/t$ amplitude was calculated to be overall dominant in all 20 simulations.

The calculated flexibility of the ribose sugar is large, and oscillations between the C2'- and C3'-endo ribose conformers can be related to the δ torsion amplitudes of $\sim 140^\circ$ and $\sim 80^\circ$, respectively. The local motions of riboses at the 3' and 5' end are not equivalent. The oscillations calculated for the $\delta 2$ torsion are harmonic, whereas the motion of ribose at the 5' end seems to be perturbed by a rotating 5'-hydroxyl. Neither the C2'- nor the C3'-endo conformer was thus calculated as dominant in nucleotides at the 3' end of the DMP molecules. On the other hand, the calculated overall motion of ribose at the 5' end was stabilized either in the C2'-endo (3dCpC, 1cCpA) or C3'-endo (#aCpC, 1cCpC, 1aCpA, CpC) conformation.

The distribution functions calculated for the glycosidic χ torsion have two maxima at $\sim 60^\circ$ and $\sim 240^\circ$, corresponding to the *syn* and *anti* sugar-to-base orientations in RNA nucleosides. The $\chi 2/anti$ conformer appearing in all of the initial structures was also calculated as being overall dominant. The average value of the $\chi 2/anti$ amplitude calculated in the ApA and CpA ($\chi 2 \sim 280^\circ$) is larger than that calculated in the ApC and CpC ($\chi 2 \sim 250^\circ$) molecules. The mean value of the $\chi 2/anti$ distributions thus specifically depends on the nucleobase type. The $\chi 1$ torsion

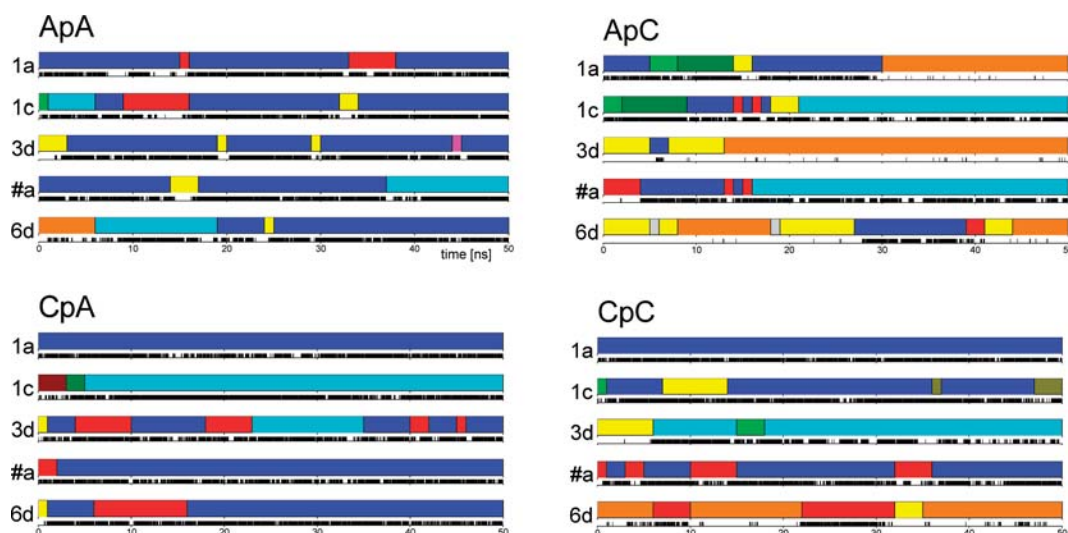


Figure 5. The correlation between the geometry of backbone segment α_2 , γ_2 , ζ_1 (upper stripe) and base stacking (lower stripe, with the color black corresponding to the geometries with stacked bases) during the time evolution calculated for the 1a, 1c, 3d, #a, and 6d initial patterns of RNA molecules. The values of the torsion angles $\alpha_2/\gamma_2/\zeta_1$ in degrees and the calculated percentage of their populations are 300/60/300 (52%, blue), 60/180/60 (19%, cyan), 60–180/60/60 (11%, orange), 60/60/60 (6%, yellow), 300/60/60 (8%, red), 300/180/300 (2%, olive), 60/180/300 (1%, green), 60/60/300 (1%, dark yellow), 60–180/180/300 (1%, wine), 60/60/180 (1%, light gray), and 300/60/180 (1%, magenta).

undergoes frequent flips between the *syn* and *anti* orientations, and the χ_1 /*syn* orientations were calculated as being overall dominant. Interestingly, the *syn* and *anti* regions of the χ_1 distributions for the ApA and ApC are smoothly jointed (via point $\chi_1 \sim 360^\circ$), unlike the separated regions calculated for the CpA and CpC molecules. The flexibility of the glycosidic bond calculated for the nucleoside at the 5' end, which is specific for adenosine and cytidine, can be again most likely attributed to the steric interactions of the 5'-hydroxyl with the nucleobases.

An Analysis of Base Stacking. The base stacking interaction plays an essential role in stabilizing the structure of nucleic acids. The model RNA molecules containing permutations of purine and pyrimidine nucleobases were synthesized to this end. They represent the simplest but still relevant models for a sequential study of base stacking in short single-strand RNA molecules.

The mutual geometrical arrangement of the nucleobases was classified according to the following criteria: the distance R_{NN} (Figure 1) between the glycosidic nitrogen atoms of the nucleobases (with these distances observed in the MD simulations being in the range of 3.3–12.0 Å), the ϕ angle between the in-plane vectors in the adenine (N9 → C6) and cytosine (N1 → C4) bases, and the ϕ angle between the bases' normal vectors. Similar geometry criteria were used for the base-stacking analysis of the nucleic acid macromolecules.⁶² In this work, the mutual inclination of the planes defined by the innering atoms of nucleobases was thus measured with the two angles (the angular geometry criteria), while their proximity was probed with the R_{NN} distance. The DMP molecules with a geometry having an $R_{NN} \leq 6.4$ Å, $\phi \leq 60^\circ$, and $|\phi| \leq 45^\circ$ (parallel bases) or $|\phi| \geq 135^\circ$ (antiparallel bases) were classified in this work as the DMP molecules with an arrangement of stacked bases. An evaluation of the base stacking was thus grounded purely on the complex geometry criteria selected to capture roughly the maximal spatial overlap of bases in the DMP molecules (for the time evolution of the three geometry parameters, see the Supporting Information). The base-stacking ratio was evaluated as the number of snapshots having a geometry corresponding to stacked bases relative to the total number of snapshots accumulated along a given trajectory (Figure 4).

The calculated base-stacking ratios depended on the starting MD structures. This ratio was for the ApC and CpC molecule in the ranges of 2–67 and 24–75%, respectively (Figure 4). The ApC and CpC simulations with the 1a and 1c initial geometries show high stacking ratios (67, 41% and 73, 75%, respectively). In these two molecules, the angular part of the geometry criteria for base-stacking classification provided more reliable measure of the stacking ratio, elsewhere frequently off the range, while the R_{NN} distance simultaneously fit the criteria for stacked bases.

The maximal base-stacking population was calculated consistently (for all five initial structures) for the ApA (64–75%) and CpA (65–77%) molecules (Figure 4). For both of these DMPs, the distance geometry criteria R_{NN} turned out to be decisive since the angular criteria appeared mostly within the target intervals. Adenosine at the 3' end of the DMP molecules thus stabilizes the stacked conformations, regardless of the nucleobase at the 5' end. The same conclusion had been drawn by Norberg and Nillson as well.⁴⁹

The Relation between the Conformation of the RNA Backbone and Base Stacking. The motions calculated for the backbone torsion angles in the DMP molecules can be roughly classified as constant (the behavior of angle β), oscillating among well-resolved amplitudes with a frequency of tens of ns (α , γ , ζ , and ϵ torsions) and oscillating fast on the nanosecond time scale between well-resolved amplitudes (δ torsion). The analysis of the conjoint motions in the RNA backbone and nucleobases thus primarily focused on the semirigid segment ζ_1 – α_2 – γ_2 .

The simultaneous population of the dominant triad/angle combination $\alpha_2/g-$, $\gamma_2/g+$, $\zeta_1/g-$ (canonical A-RNA) and the bases in stack geometry was found 72% of the time when this triad was populated. The $\alpha_2/g+$, γ_2/t , $\zeta_1/g+$ triad ranked the second most consistent with base stacking, similarly with 68%. The populations calculated for the rest of the α_2 , γ_2 , ζ_1 triads found in the MD trajectories decreased steeply, and their coherence with the base stacking was therefore overall negligible (Figure 5).

The base-stacking topology in the DMP molecules was found not only for the canonical A-RNA backbone pattern of the

TABLE 2: The Scalar Coupling Constants in Hertz Measured and Calculated in the ApA, ApC, CpA, and CpC Molecules and Their Structural Assignment

compound	unit ^a		<i>J</i> (HH)			<i>J</i> (H,P)			<i>J</i> (C,P)				<i>J</i> (H,C)	
			3',4'	4',5'	4',5''	3',P	5',P	5'',P	2',P	3',P ^b	4',P	5',P ^b	1',2/4	1',6/8
			$\delta 1/\delta 2$	$\gamma 1/\gamma 2$	$\gamma 1/\gamma 2$	$\epsilon 1$	$\beta 2$	$\beta 2$	$\epsilon 1$	$\zeta 1$	$\epsilon 1/\beta 2$	$\alpha 2$	$\chi 1/\chi 2$	$\chi 1/\chi 2$
ApA	exp. ^c	A	~5.0	2.5	3.6	~9.0			3.7	-5.4	5.3		1.9	4.2
		A	5.5	?	~3.8		?	~3.8			9.4	-5.4	2.5	3.1
	calc. ^c	A	4.0	2.4	4.9	7.2			4.5	-5.3	4.9		3.4	3.1
		A	2.8	2.1	3.2		4.2	2.3			9.6	-4.7	2.7	2.9
ApC	exp. ^c	A	6.1	2.4	3.5	8.7			3.3	-5.3	4.6		2.1	2.8
		C	?	~1.7	~2.0		~4.0	3.4			9.5	-5.1	1.4	4.5
	calc. ^c	A	6.0	2.5	4.3	5.6			3.6	-5.2	6.6		3.8	2.6
		C	3.8	2.2	4.4		2.8	3.9			9.6	-4.4	3.1	3.1
CpA	exp. ^c	C	6.8	2.6	4.0	8.7			3.4	-5.3	5.4		1.6	4.6
		A	5.5	2.6	3.0		4.3	3.8			9.3	-5.3	1.8	4.3
	calc. ^c	C	5.5	2.3	5.1	6.5			4.4	-5.3	5.4		3.3	4.0
		A	3.7	2.2	3.9		2.9	3.0			10.0	-4.7	2.6	3.5
CpC	exp. ^c	C	7.3	2.5	3.8	8.9			3.1	-5.3	6.0		1.4	4.5
		C	7.2	2.4	2.4		4.3	3.2			9.5	-5.6	1.4	4.6
	calc. ^c	C	6.4	2.3	6.0	5.6			3.2	-5.3	7.0		4.3	4.1
		C	4.5	2.3	3.4		2.9	3.0			10.0	-4.7	3.4	3.6
	exp. ^d	C	6.7	2.6	4.0	8.7			3.4	-5.3	5.5		1.7	4.6
		C	6.7	~2.5	2.7	~4.3	3.9				9.1	-5.3	1.5	4.6
	calc. ^d	C	6.0	2.3	5.3	6.1			3.6	-5.2	6.3		3.7	4.0
		C	4.4	2.3	4.3	3.2	3.5				9.6	-4.6	3.1	3.6

^a The upper and lower rows show the data for the nucleoside at the 5' and 3' end of the RNA dinucleoside, respectively. ^b The negative sign was calculated previously.⁴⁶ ^c At a temperature of 300 K. ^d At an experimental temperature of 320 K. The MD simulation at 340 K. The values were calculated only with the distributions obtained for the 1c and #a structural classes.

$\zeta 1-\alpha 2-\gamma 2$ segment but also for other patterns (Figure 5). The sequential role of purine or pyrimidine nucleobases at the 3' end of the DMP molecule discussed in the previous chapter is also critical (Figure 4). The calculations for the ApA and CpA molecules with overall high stacking ratios as well as those for the other two RNA molecules exhibited a large portion of the antiparallel arrangement of stacked bases (see the Supporting Information). Since the nucleosides at the 3' end fluctuate fairly well around the *anti* orientation of their glycosidic bond, the main reason for the antiparallel orientation of the stacked bases is a rather high population of *syn* nucleosides at the 5' end of the DMP molecules (see the Supporting Information). The analysis of the *J*-couplings assigned to the glycosidic torsion in adenosines (the following section) also supported the calculated prevalent population of the *anti* and *syn* adenosine at the 3' and 5' end, respectively.

The Structural Assignment and Calculation of the *J*-Couplings. The geometry descriptors that were used for calculating the structure of the DMP molecules are eight backbone torsion angles (two of which are correlated with the sugar pucker) and two glycosidic torsion angles (Figure 1). The *J*-couplings across the three covalent bonds (the ³*J*-couplings) were assigned to the majority of the torsion angles.^{41,46,54,55} The ²*J*-couplings near the phosphate group were assigned to the α and ζ torsion angles (see Methods Section).

The accuracy of the ideally smooth empirical dependencies for the ³*J*-couplings (Karplus equations) is limited. For example, several Karplus equations designed for one *J*-coupling may differ despite their phase factor usually remaining the same.⁴¹ Some examples of such Karplus equations were collected in our previous work (see Figure 2 of ref 46 and the references therein). Also, the one-dimensional character of the Karplus equations cannot always be guaranteed, but the effect of the neighboring geometry can be modeled.^{46,63} The absolute accuracy of the calculated *J*-couplings can thus be analyzed carefully, taking into account the mentioned facts. Being

conscious of the limitations, the theoretical method applied in this work can be used for directly probing the local geometry with NMR spectroscopy parameters.

The ³*J*(P, H5'2), ³*J*(P, H5''2), and ³*J*(P, C4'2) couplings were assigned to the $\beta 2$ torsion.^{41,46} The three Karplus equations possess different phases with respect to the calculated distribution functions centered near 180°. The dispersion of the values measured for each of the three *J*-couplings in the DMP molecules is smaller than 0.6 Hz, although their individual magnitudes are different (Table 2). This fact along with the different phase factors of the three *J*-couplings implies that the $\beta 2$ torsion should be similar in all four DMP molecules. The actual distribution function should be correspondingly single-modal, in agreement with the calculated distributions, which are centered near 180°. The absolute difference between the calculated couplings assigned to $\beta 2$ torsion and experiment was smaller than 1.5 Hz.

Interestingly, the dependence of the averaged *J*-couplings on the curvature and exact position of the central point of the $\beta 2$ distributions is large. Specifically, the center of the $\beta 2$ distribution for the ApA molecule is at 172° and has a slightly wider envelope than the distribution calculated for the ApC molecule centered at 179° (see the Supporting Information). The ³*J*(P, C4'2) couplings calculated in the ApA and ApC molecules are practically the same because the $\beta 2$ distribution and Karplus equation are "in phase", both centered near 180° (for an example of the Karplus equations, see Figure 2 in ref 46). On the other hand, the ³*J*(P, H5'2) and ³*J*(P, H5''2) couplings calculated in the ApA and ApC molecules differ by 1.4 and 1.5 Hz, respectively. Their Karplus equations are not centered at 180°, and their slopes have opposite declination in that area. The mutual phase of the Karplus equation and the corresponding distribution may thus become a very sensitive factor in the numerical averaging of *J*-couplings.

The ³*J*(H3', H4') coupling assigned to the δ torsion was used to calculate the sugar pucker. Its extreme values roughly ranged

from 0 (C2'-*endo*) to 10 Hz (C3'-*endo*) (see Figure 2 in ref 46). The measured values between 5.0 and 7.3 Hz suggest the presence of a mixture of the two conformers. Thus, a more equal population of the C2'- and C3'-*endo* pucker was measured in the ApA, whereas in the CpC molecule, the C2'-*endo* sugar dominated. The measured trends were also calculated, although the calculated values of the $^3J(\text{H3}',\text{H4}')$ couplings are overall underestimated.

The $^3J(\text{H4}',\text{H5}')$ and $^3J(\text{H4}',\text{H5}'')$ couplings assigned to the γ torsion with a trimodal pattern of 60,180, 300° have approximate magnitudes of 3, 3, 10 Hz and 3, 10, 3 Hz, respectively.⁴¹ Better overall agreement was obtained for the $^3J(\text{H4}',\text{H5}')$ coupling, with an absolute deviation of the calculated couplings from experiment that was smaller than 0.5 Hz. Systematically larger deviations from experiment (up to 2.4 Hz) were calculated for the $^3J(\text{H4}',\text{H5}'')$ coupling. The calculated γ_1 distributions show populations of its trans conformer, although $g+$ is expected to be dominant at least in the solid phase.⁵⁸ Any conclusion in that regard can be drawn only with difficulty because of the limits in accuracy as the γ_1 distributions are very narrow, which causes the numerical integration method to probe only a very confined interval of the respective Karplus equation.

The $^3J(\text{C2}'1,\text{P})$, $^3J(\text{C4}'1,\text{P})$, and $^3J(\text{H3}'1,\text{P})$ couplings were assigned to the ϵ_1 torsion.^{41,46} The two calculated $^3J(\text{C},\text{P})$ couplings deviated from the experiment by less than 1 Hz (except for the $^3J(\text{C4}'1,\text{P})$ coupling in the ApC molecule), while the deviations calculated for the $^3J(\text{H3}'1,\text{P})$ coupling ranged from 1.8 to 3.3 Hz.

The assignment of the $^2J(\text{P}, \text{C5}'2)$ and $^2J(\text{C3}'1,\text{P})$ couplings to the α_2 and ζ_1 torsion is a challenging part of this work since the dependence of the 2J -couplings on the torsion angles is less straightforward than that for the 3J -couplings. On the other hand, the ζ_1 – α_2 backbone segment suffers from a lack of relevant nuclei measurable by means of NMR spectroscopy. The two couplings therefore offer one of the limited possibilities for probing the structure of the nucleic acids near the phosphate group, as calculated previously.⁴⁶ The proposed dependence of the 2J -couplings on the ζ_1 and α_2 torsions seems to be qualitatively correct since it does not contradict the experimental data.

The $^3J(\text{H1}',\text{C2}/4)$ and $^3J(\text{H1}',\text{C6}/8)$ couplings were assigned to the χ torsion.^{55,64} The values of the J -couplings calculated by Munzarová and Sklenář for the *syn/anti* nucleosides are 5.5/2.5 Hz for both the $^3J(\text{H1}',\text{C4})$ coupling in adenosine and the $^3J(\text{H1}',\text{C2})$ coupling in cytidine, 4.0/5.0 Hz for the $^3J(\text{H1}',\text{C8})$ coupling in adenosine, and 5.5/4.5 Hz for the $^3J(\text{H1}',\text{C6})$ coupling in cytidine.⁶⁴ The *syn* and *anti* regions of both dependencies used for the determination of the χ torsion actually coincided with the two maxima on the Karplus curves.^{41,64}

The $^3J(\text{H1}',\text{C8})$ couplings measured in the different adenosines have different values, namely, ApA (4.2, 3.1 Hz), ApC (2.8 Hz), and CpA (4.3 Hz). A similar variation was also measured also for the $^3J(\text{H1}',\text{C4})$ couplings (see Table 2). On the contrary, both couplings assigned to the χ torsion measured in cytidines varied by less than 0.2 Hz.

The calculated χ distributions are sequence-dependent. The χ_1 and χ_2 distributions both have a similar shape, irrespective of the nucleobases. Specifically, the χ_1 distributions possess a sharp *syn* peak as compared to the χ_2 distributions, and χ_2 distributions have a broad *anti* peak; this, however, contradicts the experimental data. The $^3J(\text{H1}',\text{C2})$ and $^3J(\text{H1}',\text{C6})$ couplings measured in cytidines have practically the same values, while the calculated ones were modulated by the specificity of the

χ_1 , χ_2 distributions (Table 2). This discrepancy can be most likely attributed to the drawbacks of the force field used.

The $^3J(\text{H1}',\text{C2})$ couplings measured in cytidines (~ 1.5 Hz) are certainly below the proposed *syn* (5.5 Hz) and *anti* (2.5 Hz) magnitudes.⁶⁴ Therefore, glycosidic motion in all cytidines should be rather uniform and delocalized, probably corresponding to the broad distribution that is dominant in the *anti* region.

The increase of the $^3J(\text{H1}',\text{C8})$ coupling and simultaneous decrease of the $^3J(\text{H1}',\text{C4})$ coupling as measured in adenosines can be assigned to the increase of the *anti* population and vice versa.^{55,64} (The global maxima of the dependencies calculated for the two J -couplings swap: $^3J(\text{H1}',\text{C8})/\textit{syn} < ^3J(\text{H1}',\text{C8})/\textit{anti}$ and $^3J(\text{H1}',\text{C4})/\textit{syn} > ^3J(\text{H1}',\text{C4})/\textit{anti}$.) The *anti* population of the 3'-adenosine in the ApA molecule should thus be higher than the *anti* population of the 5'-adenosine. Similarly, the *syn* population of adenosine in the ApC should be higher than the *syn* population of adenosine in the CpA molecule. We note that this site-specific assignment for the glycosidic torsion of adenosines in the DMP molecules is in agreement with the calculated stacking ratios (see Figure 4). The absolute deviation of the calculated J -couplings assigned to the χ torsion in adenosines from the experiment ranged up to 1.7 Hz.

J -Couplings for the CpC Molecule at a Temperature of 340 K. The MD calculations at a temperature of 340 K were carried out only for the CpC molecule with the 1c and #a initial structure (see the Supporting Information). The dynamics of the CpC molecule at the higher temperature possess more frequent oscillations of the torsion angles accompanied with a decrease of the base-stacking ratio from 62 to 46% for the 1cCpC calculation and from 73 to 57% for the #aCpC calculation. A similar drop in the base-stacking ratio had previously been calculated in the ApA dinucleoside by Norberg.⁶⁵ The calculated distributions are quite strongly modulated by temperature increase (Figure 3).

The absolute shift of the measured J -couplings with the temperature increase was smaller than 0.7 Hz as measured for the $^3J(\text{P}, \text{H5}'2)$ coupling assigned to the β_2 torsion. The decrease of the $^3J(\text{H3}',\text{H4}')$ coupling by 0.5 Hz measured in both cytidines can be assigned to the more equal populations of the C2'- and C3'-*endo* sugar pucker conformers when compared to the populations at lower temperature (Table 2). (The equilibrium can be estimated for the J -coupling magnitude of ~ 5 Hz.)

The faster molecular motions calculated for the backbone atoms related to the torsion angles modulate the shape of the calculated distributions (Figure 3). The J -couplings calculated at the higher temperature assigned to the backbone torsions mostly follow the measured trends, except for the $^3J(\text{H3}'1,\text{P})$ and $^3J(\text{P}, \text{H5}'2)$ couplings (Table 2). The absolute temperature shift of the calculated J -couplings was smaller than 0.9 Hz. Interestingly, the temperature shift calculated for the J -couplings assigned to the glycosidic torsion ranged from 0.3 to 1.5 Hz. This fact again pointed to possible deficiencies in the dynamics calculated for the glycosidic motions and to the fact that a higher overall reliability should be attributed to the results obtained for the backbone torsion angles.

Conclusion

The NMR scalar J -coupling constants measured in the ApA, ApC, CpA, and CpC RNA dinucleoside monophosphates were assigned to the α , β , γ , δ , ϵ , and ζ backbone torsion angles and to the χ glycosidic torsion. Assembling the local structural information for the individual torsion angles that had been validated with the NMR J -couplings allows for the determination

of the individual global structural and dynamical features of the RNA molecules.

The measured scalar J -couplings were correlated with the J -couplings calculated using the Karplus equations averaged with the distributions for the torsion angles obtained from the MD simulations. The sampling of the structural patterns in the MD calculations was enhanced by taking the initial patterns populated in rRNA resolved in the solid state into account.^{57,58} The calculation method can thus reliably and consistently probe the dynamical multi-amplitude character of the torsion angles and their relative populations with NMR spectroscopy.

Overall, good agreement of the measured and calculated J -couplings was obtained particularly for the backbone torsions. The absolute average deviation from experiment was 0.4, 0.6, 1.6, 1.0, and 1.1 Hz for the ${}^2J(\text{P,C})/(\alpha,\zeta)$, ${}^3J(\text{P,C})/(\varepsilon,\beta)$, ${}^3J(\text{P,H})/(\varepsilon,\beta)$, ${}^3J(\text{H,H})/(\delta,\gamma)$, and ${}^3J(\text{C,H})/(\chi)$ couplings. For the calculation of the α and ζ torsions, the ${}^2J(\text{P,C})$ scalar couplings were newly used. The calculated scalar J -couplings assigned to the glycosidic torsion contradict the measured data, most likely because of the possible drawbacks of the force field used, which probably failed in the qualitative prediction of this large-amplitude motion in the nucleosides of the four RNA molecules. In that respect, a better robustness of the calculation method should be expected for the relatively rigid geometry as calculated for the RNA backbone. However, the "static assignment" of the experimental J -couplings across the glycosidic torsion (i.e., the direct relation of the measured J -couplings to the χ torsion angles via Karplus curves without introducing the distributions for torsion angles) implied different dynamics in cytidines (steady motion) and adenosines (sequence-specific motion) in the DMP molecules.

The magnitudes of the calculated J -couplings depend critically on the distribution functions of the torsion angles. Another critical parameter is the mutual symmetry of the distribution and Karplus equation, as was shown for the three 3J -couplings assigned to the β torsion.

The set of distribution functions for the torsion angles, which is coherent with the structural pattern of A-RNA, was overall resolved as the most populated. Specifically, the distributions calculated for the α_2 , γ_2 , ζ_1 torsion angles with the pattern of A-RNA dominate overall, and such geometries enhance the stacking of the bases in the RNA dinucleoside monophosphates. Further, an important prerequisite for the base-stacking topologies validated in this study with the NMR data is the presence of adenosine (the purine nucleobase) at the 3' end of the RNA dinucleoside monophosphates. The ApA and CpA dinucleosides thus possess a higher stacking ratio than the ApC or CpC ones. The calculated base-stacking ratio in the CpC molecule decreases as the temperature increases.

Acknowledgment. This work was supported by the Academy of Sciences of the Czech Republic, Grant Nos. IAA400550701, AV0Z50040507, AV0Z50520701, AV0Z40550506, and AV0Z50040702, and by the Grant Agency of Charles University, Grant GAUK No. 58708. V.S. was supported by a Human Frontier Science Program (HFSP) Young Investigator's Grant.

Supporting Information Available: Calculated trajectories and distribution functions for the α , β , γ , δ , ε , ζ , and χ torsion angles, trajectories of the geometry parameters for analysis of base stacking, and an explicit form of the Karplus equations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147.
- (2) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *Cold Spring Harbor Monograph Series*, 2006.
- (3) Fire, A.; Xu, S. Q.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. *Nature* **1998**, *391*, 806.
- (4) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
- (5) Wimberly, B. T.; Brodersen, D. E., Jr.; Morgan-Warren, R.; Carter, A. P.; Vornheim, C.; Hartsch, T.; Ramakrishnan, V. *Nature* **2000**, *407*, 327.
- (6) Schuwirth, B. S.; Borovinskaya, M. A.; Hau, C. W.; Zhang, W.; Vila-Sanjurjo, A.; Holton, J. M.; Cate, J. H. D. *Science* **2005**, *310*, 827.
- (7) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935.
- (8) Harms, J.; Schluenzen, F.; Zarivach, R.; Bashan, A.; Gat, S.; Agmon, I.; Bartels, H.; Franceschi, F.; Yonath, A. *Cell* **2001**, *107*, 679.
- (9) Korostelev, A.; Trakhanov, S.; Laurberg, M.; Noller, H. F. *Cell* **2006**, *126*, 1065.
- (10) Tanaka, Y.; Taira, K. *Chem. Commun.* **2005**, 2069.
- (11) Tanaka, Y.; Kasai, Y.; Mochizuki, S.; Wakisaka, A.; Morita, E. H.; Kojima, C.; Toyozawa, A.; Kondo, Y.; Taki, M.; Takagi, Y.; Inoue, A.; Yamasaki, K.; Taira, K. *J. Am. Chem. Soc.* **2004**, *126*, 744.
- (12) Frank, J.; Agrawal, R. K. *Nature* **2000**, *406*, 318.
- (13) Frank, J.; Spahn, C. M. T. *Rep. Prog. Phys.* **2006**, *69*, 1383.
- (14) Mitra, K.; Frank, J. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 299.
- (15) Li, P. T. X.; Viereg, J.; Tinoco, I. *Annu. Rev. Biochem.* **2008**, *77*, 77.
- (16) Spirin, A. S. *FEBS Lett.* **2002**, *514*, 2.
- (17) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911.
- (18) Mathews, D. H.; Turner, D. H. *Curr. Opin. Struct. Biol.* **2006**, *16*, 270.
- (19) Leontis, N.; Stombaugh, J.; Westhof, E. *Biochimie* **2002**, *84*, 961.
- (20) Lescoute, A.; Leontis, N. B.; Massire, C.; Westhof, E. *Nucleic Acids Res.* **2005**, *33*, 2395.
- (21) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215.
- (22) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 14.
- (23) Nissen, P.; Ippolito, J. A.; Ban, N.; Moore, P. B.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4899.
- (24) Lescoute, A.; Westhof, E. *RNA* **2006**, *12*, 83.
- (25) Kolk, M. H.; van der Graaf, M.; Wijmenga, S. S.; Pleij, C. W. A.; Heus, H. A.; Hilbers, C. W. *Science* **1998**, *280*, 434.
- (26) Ferner, J.; Villa, A.; Duchardt, E.; Widjajakusuma, E.; Wohnert, J.; Stock, G.; Schwalbe, H. *Nucleic Acids Res.* **2008**, *36*, 1928.
- (27) Schwalbe, H.; Buck, J.; Furtig, B.; Noeske, J.; Wohnert, J. *Angew. Chem., Int. Ed.* **2007**, *46*, 1212.
- (28) Varani, G.; Aboulela, F.; Allain, F. H. T. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51.
- (29) Getz, M.; Sun, X. Y.; Casiano-Negroni, A.; Zhang, Q.; Al-Hashimi, H. M. *Biopolymers* **2007**, *86*, 384.
- (30) Flinders, J.; Dieckmann, T. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 137.
- (31) Tanaka, Y.; Ono, A. *Dalton Trans.* **2008**, 4965.
- (32) Li, W.; Frank, J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16540.
- (33) Rhodes, M. M.; Reblova, K.; Sponer, J.; Walter, N. G. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13380.
- (34) Rázga, F.; Koča, J.; Mokdad, A.; Šponer, J. *Nucleic Acids Res.* **2007**, *35*, 4007.
- (35) McDowell, S. E.; Spackova, N.; Sponer, J.; Walter, N. G. *Biopolymers* **2007**, *85*, 169.
- (36) Auffinger, P.; Hashem, Y. *Curr. Opin. Struct. Biol.* **2007**, *17*, 325.
- (37) Qin, P. Z.; Dieckmann, T. *Curr. Opin. Struct. Biol.* **2004**, *14*, 350.
- (38) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564.
- (39) Furtig, B.; Richter, C.; Wohnert, J.; Schwalbe, H. *ChemBioChem* **2003**, *4*, 936.
- (40) Přechtělová, J.; Padrt, P.; Munzarová, M. L.; Sklenář, V. *J. Phys. Chem. B* **2008**, *112*, 3470.
- (41) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287.
- (42) Grzesiek, S.; Barfield, M.; Cordier, F.; Dingley, A.; Feigon, J.; Nicholson, L. K. *Abstr. Pap. Am. Chem. Soc.* **2000**, *219*, U139.
- (43) Grzesiek, S.; Cordier, F.; Jaravine, V.; Barfield, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *45*, 275.
- (44) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.
- (45) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.

- (46) Sychrovský, V.; Vokáčová, Z.; Šponer, J.; Špačková, N.; Schneider, B. *J. Phys. Chem. B* **2006**, *110*, 22894.
- (47) Rinnenthal, J.; Richter, C.; Ferner, J.; Duchardt, E.; Schwalbe, H. *J. Biomol. NMR* **2007**, *39*, 17.
- (48) Ravindranathan, S.; Kim, C. H.; Bodenhausen, G. *J. Biomol. NMR* **2003**, *27*, 365.
- (49) Norberg, J.; Nilsson, L. *J. Am. Chem. Soc.* **1995**, *117*, 10832.
- (50) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, CA, 2004.
- (51) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
- (52) Marino, J. P.; Schwalbe, H.; Griesinger, C. *Acc. Chem. Res.* **1999**, *32*, 614.
- (53) Yokoyama, S.; Inagaki, F.; Miyazawa, T. *Biochemistry* **1981**, *20*, 2981.
- (54) Haasnoot, C. A. G.; Deleeuw, F.; Altona, C. *Tetrahedron* **1980**, *36*, 2783.
- (55) Munzarová, M. L.; Sklenář, V. *J. Am. Chem. Soc.* **2003**, *125*, 3649.
- (56) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751.
- (57) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* **2008**, *14*, 465.
- (58) Schneider, B.; Morávek, Z.; Berman, H. M. *Nucleic Acids Res.* **2004**, *32*, 1666.
- (59) Svozil, D.; Kalina, J.; Omelka, M.; Schneider, B. *Nucleic Acids Res.* **2008**, *36*, 3690.
- (60) Krasovská, M. V.; Šefčíková, J.; Réblová, K.; Schneider, B.; Walter, N. G.; Šponer, J. *Biophys. J.* **2006**, *91*, 626.
- (61) Réblová, K.; Lankaš, F.; Rázga, F.; Krasovská, M. V.; Koča, J.; Šponer, J. *Biopolymers* **2006**, *82*, 504.
- (62) Gabb, H. A.; Sanghani, S. R.; Robert, C. H.; Prévost, C. *J. Mol. Graphics* **1996**, *14*, 6.
- (63) Sychrovský, V.; Muller, N.; Schneider, B.; Smrecki, V.; Špirko, V.; Šponer, J.; Trantírek, L. *J. Am. Chem. Soc.* **2005**, *127*, 14663.
- (64) Munzarová, M. L.; Sklenář, V. *J. Am. Chem. Soc.* **2002**, *124*, 10666.
- (65) Norberg, J.; Nilsson, L. *J. Phys. Chem.* **1995**, *99*, 13056.

JP809762B

Structural Interpretation of J Coupling Constants in Guanosine and Deoxyguanosine: Modeling the Effects of Sugar Pucker, Backbone Conformation, and Base Pairing

Zuzana Vokáčová,^{*,†} F. Matthias Bickelhaupt,^{*,‡} Jiří Šponer,[§] and Vladimír Sychrovský^{*,†}

Institute of Organic Chemistry and Biochemistry, v.v.i., Academy of Sciences of the Czech Republic, Flemingovo Square 2, 166 10 Prague 6, Czech Republic, Theoretical Chemistry and Amsterdam Center for Multiscale Modeling, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands, and Institute of Biophysics, v.v.i., Academy of Sciences of the Czech Republic, Královopolská 135, 612 65 Brno, Czech Republic

Received: March 19, 2009; Revised Manuscript Received: May 20, 2009

The $^3J(\text{C8}–\text{H1}')$, $^3J(\text{C4}–\text{H1}')$, $^1J(\text{C8}–\text{H8})$, $^1J(\text{C1}'–\text{H1}')$, $^1J(\text{C2}'–\text{H2}')$, and $^1J(\text{C2}'–\text{H2}''')$ indirect scalar coupling constants were calculated with the density functional theory in the deoxyguanosine and riboguanosine molecules. The following geometry descriptors were considered in analysis of the structural dependence of the six J couplings: the glycosidic torsion angle χ and conformation of the hydroxymethyl group at the C4' carbon of sugar mimicking the backbone residue and the sugar pucker (C2'-, C3'-endo). The $^3J(\text{C8}–\text{H1}')$ and $^3J(\text{C4}–\text{H1}')$ couplings, which are typically assigned to the χ torsion, also depended on the sugar pucker, although the calculated dependence of the latter coupling on sugar pucker was nearly negligible. New parametrization of the Karplus equations, taking into account the stereoinversion effect at the glycosidic nitrogen atom and solvent effects, was calculated for the $^3J(\text{C8}–\text{H1}')$ and $^3J(\text{C4}–\text{H1}')$ coupling assigned to the χ torsion. The calculated phase shift of χ torsion angle in these new Karplus equations was larger by $\sim 10^\circ$ compared to its commonly accepted value of 60° (Wijmenga, S. S.; van Buuren, B. N. M. *Prog. NMR Spectrosc.* **1998**, *32*, 287.). The calculated $^1J(\text{C2}'–\text{H2}')$ and $^1J(\text{C2}'–\text{H2}''')$ coupling dominantly depended on the sugar type (deoxyribose or ribose) and its pucker, while the $^1J(\text{C1}'–\text{H1}')$ and $^1J(\text{C8}–\text{H8})$ coupling dominantly depended on the glycosidic torsion angle, although quantitatively, all four 1J couplings depended on both geometry parameters. The dependences of j -couplings on the torsion angle χ calculated in isolated nucleosides were compared with those taking into account the effect of base pairing occurring in the WC/SE RNA base pair family, which appeared to be minor. The calculated 3J couplings agreed well with available experimental data similarly as the 1J couplings, although lack of experimental data diminished more reliable validation of the later couplings.

1. Introduction

Biochemical properties and biological functions of the nucleic acids (NAs) are strongly related to their structural dynamics. One of the relevant possibilities of how NAs can be studied in aqueous solution is offered by NMR spectroscopy.^{1,2}

The nucleoside unit is a basic building block of NAs. The measurements of J couplings in NAs can provide local information about the geometry parameters between coupled nuclei. The three-bond J couplings (3J couplings) are typically used for determining the torsion angles in both NAs and peptides. For example, the $^3J(\text{C8}–\text{H1}')$ and $^3J(\text{C4}–\text{H1}')$ couplings are assigned to the glycosidic torsion angle in nucleosides. However, also, other kinds of J couplings can be used for determining the geometry of NA molecules. For example, the 2J and 4J couplings, which were recently calculated in structural patterns of rRNA, showed their dependences on magnitudes of the NA backbone torsion angles.³ Complex analysis of the sugar-to-base orientation in nucleosides with NMR can provide not only

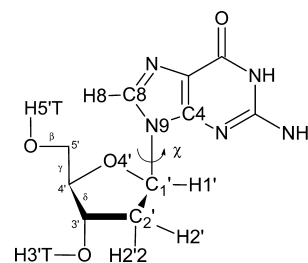


Figure 1. Sketch of the deoxyguanosine molecule with numbering of atoms and torsion angles; $\chi = \text{O4}'–\text{C1}'–\text{N9}–\text{C4}'$, $\chi' = \text{H1}'–\text{C1}'–\text{N9}–\text{C8} + 60^\circ$, $\beta = \text{H5'T}–\text{O5}'–\text{C5}'–\text{C4}'$, $\gamma = \text{O5}'–\text{C5}'–\text{C4}'–\text{C3}'$, $\delta = \text{C5}'–\text{C4}'–\text{C3}'–\text{O3}'$.

important local structural information, but it can also substantially help to refine the global topologies of NAs and to link the structure with its dynamics. Both types of sugar appearing in the deoxyguanosine (dG) and guanosine (rG), the respective DNA and RNA nucleoside, were investigated in this work. Their main geometry parameters (Figure 1) are the glycosidic torsion angle χ and the sugar pucker conformation (Figure 2). It is well established that RNA with ribose prefers the C3'-endo conformation as dominant, while DNA with deoxyribose mostly populates the C2'-endo arrangements. However, in noncanonical regions of functional RNA molecules C2'-endo is a rather frequent minor sugar pucker conformation that is essential for

* To whom correspondence should be addressed. E-mails: zuzana.vokacova@uochb.cas.cz (Z.V.), FM.Bickelhaupt@few.vu.nl (F.M.B.), vladimir.sychrovsky@uochb.cas.cz (V.S.).

[†] Institute of Organic Chemistry and Biochemistry, v.v.i., Academy of Sciences of the Czech Republic.

[‡] Scheikundig Laboratorium der Vrije Universiteit.

[§] Institute of Biophysics, v.v.i., Academy of Sciences of the Czech Republic.

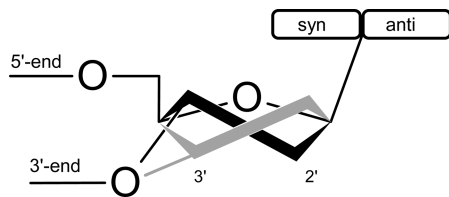


Figure 2. Schematic representation of the C2'-endo (gray) and C3'-endo (black) sugar pucker corresponding to the South and North conformation, respectively, and definition of the syn ($50^\circ < \chi < 80^\circ$) and anti ($180^\circ < \chi < 280^\circ$) regions of χ torsion, describing different orientations of the nitrogenous base with respect to sugar.

important noncanonical segments and RNA building blocks,^{4,5} while, for example, A-DNA and A to B intermediate DNA structures are characterized by A-like C3'-endo puckers.^{6,7} The effect of the sugar puckering and the sugar-to-base orientation on NMR parameters has been studied extensively both experimentally^{8–14} and theoretically.^{1,15–19}

A rather basal prerequisite for an accurate structural interpretation of NMR experimental data is to describe correctly all effects which can influence the magnitudes of measured NMR parameters. Obviously, one needs to first know the basic dependence on the assigned geometry parameter. However, this primary dependence can be substantially modulated by other structural parameters due to the natural complex flexibility of nucleic acids.^{20,21} Also, proper inclusion of solvent seems to be essential for accurate calculations of the NMR parameters.¹⁷

In this work, we address the issue of accuracy of the fitted Karplus equations using comprehensive modeling of both the local geometry and solvent effects. The empirical Karplus equations derived with the experimental data commonly suffer from an insufficient number of well-resolved points used for their fitting. The quality of the fit depends crucially on the number of available experimental data points (typically the 3J coupling/torsion angle). Such benchmark cases unfortunately offer rather limited variation in the structural coordinate since they mostly correspond to molecules with a multiply validated geometry like the Dickerson dodecamer, in order to prevent misinterpretation of the experimental data. Further, a priori known torsion angles which can be used for the fitting do not have to sufficiently characterize all important conformational regions. Typical well-established classes of NAs with well-defined geometries like A-, B-, or Z-DNA and A-RNA with canonical base pairs offer only narrow ranges of χ torsion within the syn or anti region (Figure 2).¹⁸ On the other hand, especially RNA molecules offer astonishing variability of local structures as well as global folding topologies associated with wide ranges of dynamical behavior where sampling of a more complex distribution of the glycosidic torsions can be expected. Theoretical parametrizations of the Karplus curves, which take into account all important geometry parameters in a complex way, thus can improve their accuracy and reliability. Theory can further model the trends in J couplings by solvation or different solute–solvent interactions, providing thus an estimate of robustness for the calculated dependencies of NMR parameters.

A comprehensive computational study of the six J couplings in the dG and rG nucleosides carried out in this study was focused on the following structural features and effects: (i) variation of the glycosidic torsion angle χ , (ii) effect of the B- and Z-DNA backbone conformations considered for the hydroxymethyl group at the C4' carbon of the sugar moiety, (iii) effect of sugar specificity in the deoxyribose (dG) and ribose (rG) nucleosides, (iv) effect of sugar pucker (C2'-, C3'-endo), (v) effect of solvent (gas-phase versus implicit model of water

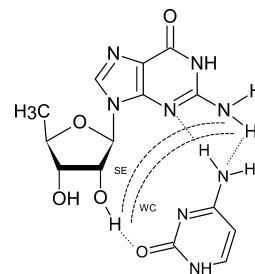


Figure 3. Sketch of the Watson–Crick/sugar edge (WC/SE) cytosine/guanine base pair including putative noncovalent interactions (dotted lines).

solvent), and (vi) effect of base pairing in non-Watson–Crick RNA base pairs (Figure 3). The later task is motivated by the fact that folded RNA molecules contain a wide range of functionally important non-Watson–Crick base pairs (altogether, six geometrical families) that directly include the ribose via its 2'-hydroxyl group in base pairing.²² The noncanonical base pairs are of primary importance in building up key functional RNA building blocks,²³ while their direct identification by NMR techniques is not straightforward.²⁴

2. Methods Section

All geometries were optimized with the B3LYP functional and the 6-31G** basis set. The glycosidic torsion χ was varied stepwise as the main geometry parameter (the χ scans). The other geometry parameters were either kept fixed in each step (backbone, see below) or relaxed (sugar pucker and all remaining geometry parameters), thus giving rise to separate χ scans (Supporting Information).

The hydroxymethyl group at the C4' carbon of sugar modeling of the part of NA backbone was kept fixed in the geometry corresponding to B-DNA ($\beta = 176^\circ$, $\gamma = 48^\circ$, Figure 1), Z-DNA ($\beta = 183^\circ$, $\gamma = 179^\circ$), and A-RNA ($\beta = 173^\circ$, $\gamma = 54^\circ$). Two separate χ scans with the C2'- or C3'-endo sugar pucker (Figure 2) were carried out for all dG and rG nucleosides. We note that the sugar moiety preserved its starting conformation C2'- or C3'-endo obtained for the global minima structure and did not interconvert when the χ torsion was smoothly varying. The geometries of the B-DNA dG nucleosides were obtained previously²⁰ as well as the geometries of the WC/SE base pairs.^{25,26}

The NMR scalar spin–spin coupling constants^{27,28} were computed using the coupled perturbed density functional theory (CP-DFT) method^{29,30} with the B3LYP functional, which was found to possess the best overall performance among other commonly used DFT functionals,^{29,30} and the IgloIII basis set³¹ by including all four J coupling terms (DSO, PSO, FC, SD). The $^nJ(X,Y)$ couplings in units of Hz were calculated between atoms X and Y (actually the ^1H or ^{13}C isotopes) separated by n bonds.

The geometries were optimized in the gas phase, and J coupling were computed both in the gas phase and with the polarized continuum model of water (PCM)³² added subsequently for the gas-phase equilibrium geometries.

The effect of atomic basis set size on the magnitude of the J couplings was calculated for the B-DNA guanosine with C2'-endo sugar ($\chi = 239.1^\circ$, global energy minimum). The difference between J couplings calculated with the Iglo II and Iglo III bases was smaller than 0.1 and 1.8 Hz for the 3J and 1J couplings, respectively. This basis set effect corresponding roughly to the relative error of 1% is in agreement with our previous study.³

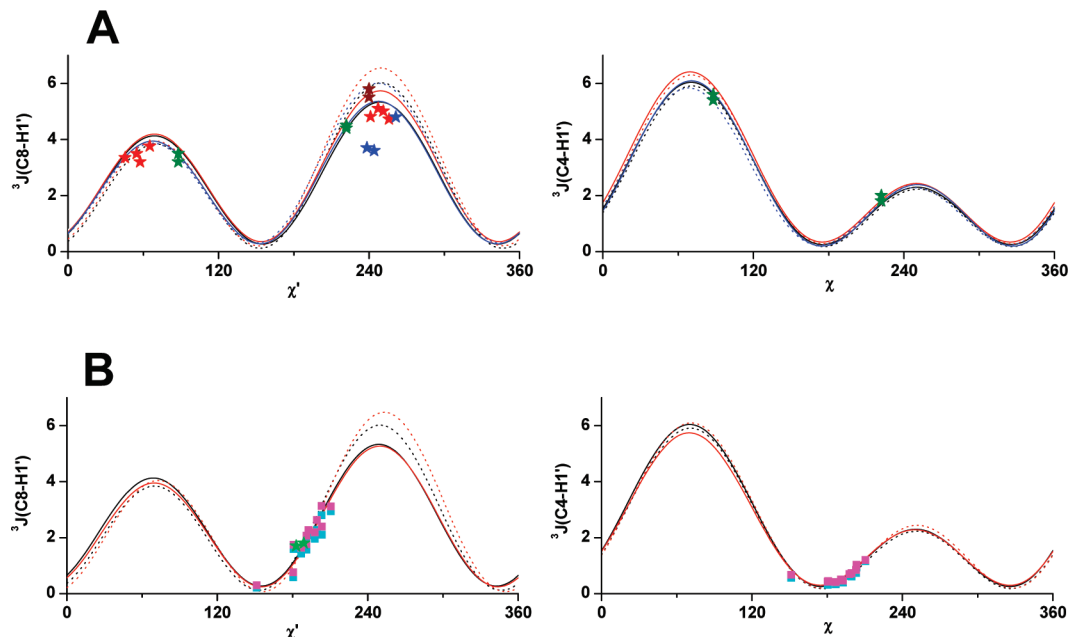


Figure 4. The fitted Karplus equations for the $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ coupling constants in Hz plotted as function of the χ' ($\chi' = \text{H1}'-\text{C1}'-\text{N9}-\text{C8} + 60^\circ$)¹⁶ and the χ (the glycosidic torsion) angle in degrees, respectively. The C2'- and C3'-endo sugar pucker conformations correspond to the solid and dotted lines, respectively. (A) The dG nucleosides; B-DNA without solvent (black), B-DNA including PCM solvent (red), and Z-DNA without solvent (blue). (B) The rG nucleosides; B-DNA without solvent added for comparison (black), A-RNA without solvent (red), and WC/SE base pairs without and including water solvent (blue and magenta squares, respectively). The experimental 3J couplings in DNA (red,¹⁸ olive,³⁴ blue,¹⁴ and brown³⁵ asterisk) and RNA (green asterisk¹²).

All calculations were done with the Gaussian 03 program package.³³

All calculated J couplings are listed in the Supporting Information (Tables S1–S10).

3. Results and Discussion

The three-bond couplings $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ across the glycosidic linkage are usually assigned to the χ torsion. In this work, we analyzed the finer effects of the backbone conformation, composition of sugar, and its pucker, base pairing, and solvent on the primary dependence of the six J coupling constants on the glycosidic torsion angle χ in the dG and rG nucleosides. The three-bond couplings $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ across the glycosidic linkage are typically assigned to the χ torsion. Such structural dependence of the one-bond couplings $^1J(\text{C2}'-\text{H2}')$, $^1J(\text{C2}'-\text{H2}'')$, and $^1J(\text{C8-H8})$ in NA nucleosides was, to our best knowledge, studied in this work for the first time. Dependence of the two 3J couplings and the $^1J(\text{C1}'-\text{H1}')$ coupling on the χ torsion only was studied previously in the DNA nucleosides.¹⁶ By including also the other 1J couplings in this study focused on determination of the χ torsion, we would like to expand our knowledge about their dependence on this basic structural descriptor in NA nucleosides.

3.1. The 3J Couplings. Dependence of the $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings on the glycosidic torsion is large and actually dominates among the dependencies on other geometry parameters, as has been shown in the earlier studies.^{1,15,16} This can be expected since the spin–spin coupling pathway of the two 3J couplings and atoms defining the χ torsion mostly coincide. Munzarová et al.¹⁶ recently suggested new theoretical Karplus equations for the two 3J couplings which were specific with regard to the kind of base in the DNA nucleosides.¹⁶ The effect of sugar pucker on the magnitude of the $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings (smaller than 0.8 Hz) was reported just for two geometries corresponding to the energy minima in the syn and anti region.¹⁶

Before we proceed to discussion on the calculated 3J couplings, we comment on their individual assignment to torsion angles. The $^3J(\text{C4-H1}')$ coupling was assigned to the χ torsion, although the atoms defining χ torsion ($\text{O4}'-\text{C1}'-\text{N9}-\text{C4}$) are not the same as the spin–spin pathway of the $^3J(\text{C4-H1}')$ coupling (Figure 1). The $^3J(\text{C8-H1}')$ coupling requires special assignment to the appropriate torsion angle. We used for this purpose the χ' torsion; this torsion equals χ only upon exact nucleobase planarity at nitrogen N9 (Figure 1). The same assignment was actually used previously by Munzarová et al.¹⁶ (See also the section 3.2.)

The calculated dependencies of the two 3J couplings on the χ - and χ' torsion angles show basically the same shape, but their calculated dependence on the sugar pucker differs (Figure 4). The calculated profile of the $^3J(\text{C4-H1}')$ coupling was less dependent on the conformation of sugar than the $^3J(\text{C8-H1}')$ one, in agreement with the calculations by Munzarová.¹⁶ The changes in magnitudes of both 3J couplings due to different sugar pucker were overall smaller than 0.4 and 0.8 Hz for the $^3J(\text{C4-H1}')$ and $^3J(\text{C8-H1}')$ coupling, respectively. Both 3J couplings calculated within the syn region of χ torsion in the C3'-endo B-DNA nucleosides are smaller compared to the C2'-endo ones, while the opposite trend was calculated in the anti region, again in agreement with Munzarová's study.¹⁶ The sugar pucker effect on both 3J couplings was negligible when χ torsion was in between the syn and anti regions. The same trend was actually obtained in the Z-DNA dG nucleosides and also when the PCM model of water solvent was applied (Figure 4).

Application of the PCM solvent led to the increase of the two calculated 3J couplings independent of sugar pucker, which ranged from 0.1 to 0.4 Hz. The calculated trends were the same in both dG and rG nucleosides.

Surprisingly, the base pairing interactions in the direct and water-mediated *cis*-WC/SE G/rG base pairs (selected as a representative example of base pairs interacting via their sugar edges) did not modify the trends calculated for isolated

nucleosides (Figures 4B and 6). The 3J couplings calculated in the WC/SE base pairs nicely fit the trends shown by the fitted Karplus equations, although the sugar edge is relatively close to the glycoside bond of the rG nucleoside.

Available experimental data^{12,14,18,34,35} agree well with the calculated dependencies (Figure 4). Probably the best agreement with experiment was achieved for the $^3J(\text{C8}-\text{H1}')$ and $^3J(\text{C4}-\text{H1}')$ couplings measured in DNA G quartet.^{10,34} To our best knowledge, the only experimental value of the $^3J(\text{C8}-\text{H1}')$ couplings measured in RNA by Hines et al.¹² perfectly agrees with the calculated coupling (Figure 4B). The $^3J(\text{C8}-\text{H1}')$ (4.4 ± 0.1 Hz) and $^3J(\text{C4}-\text{H1}')$ (2.8 ± 0.4 Hz) couplings measured in RNA by Schwalbe et al.³⁶ could not be correlated with molecular structure since the torsion in a mononucleoside shows conformational averaging; nevertheless, the measured ranges basically agree with our calculations.

3.2. Parametrization of Karplus Equations for the 3J Couplings. The mathematical form of how the 3J couplings depend on the corresponding torsion angle suggested first by Karplus^{37,38}

$$^3J(\text{Cn} - \text{H1}') = A \cos^2(\chi - \varphi) + B \cos(\chi - \varphi) + C \quad (1)$$

was also adopted here for the $^3J(\text{Cn}-\text{H1}')$, $n = 4, 8$, couplings. The amplitudes A and B and parameter C are normally fitted on the basis of measurements (the empirical Karplus equation), while the phase shift φ is usually considered constant. Alternatively, they can be fitted for the calculated data points. The magnitude of the φ shift corresponds here to the difference between the χ torsion and the actual torsion angle probed with J coupling, which is given by the three bonds interconnecting the $\text{Cn}-\text{H1}'$, $n = 4, 8$, atoms. In this case, the χ torsion ($\text{O4}'-\text{C1}'-\text{N9}-\text{C4}$) differs from the spin-spin pathway of the $^3J(\text{C4}-\text{H1}')$ and $^3J(\text{C8}-\text{H1}')$ couplings by about 60° and 120° , respectively.

The accuracy of the torsion angles resolved in practice with measured 3J couplings depends on the quality of the fitted Karplus curve. The validity and robustness of the fit with respect to complex geometry parameters or solute-solvent interactions is therefore worthwhile to model since such effects may lower the quality of the outgoing NMR structural data.

As was mentioned, the φ phase shift in empirical Karplus equations is usually fixed to a certain value, which can be expected from the steric arrangement of atoms and corresponds to the difference between the assigned and “NMR probed” torsion angle (Ipel,^{9,10} Wijmenga,¹ Trantírek;¹⁸ eqs 2–5). However, Munzarová et al.¹⁶ pointed out that the φ phase shift of the $^3J(\text{C4}-\text{H1}')$ coupling assigned to the χ torsion actually deviates from the ideal magnitude of 60° when included in the fitting procedure (eqs 6 and 7).

$$^3J_{\text{Wijmenga}}(\text{C4} - \text{H1}') = 4.7 \cos^2(\chi - 60^\circ) + 2.3 \cos(\chi - 60^\circ) + 0.1 \quad (2)$$

$$^3J_{\text{Wijmenga}}(\text{C8} - \text{H1}') = 4.5 \cos^2(\chi - 60^\circ) - 0.6 \cos(\chi - 60^\circ) + 0.1 \quad (3)$$

$$^3J_{\text{Trantírek}}(\text{C4} - \text{H1}') = 4.4 \cos^2(\chi - 60^\circ) + 1.4 \cos(\chi - 60^\circ) + 0.1 \quad (4)$$

$$^3J_{\text{Trantírek}}(\text{C8} - \text{H1}') = 4.1 \cos^2(\chi - 60^\circ) - 0.7 \cos(\chi - 60^\circ) + 0.1 \quad (5)$$

$$^3J_{\text{Munzarová}}(\text{C4} - \text{H1}') = 3.6 \cos^2(\chi - 68.6^\circ) + 1.8 \cos(\chi - 68.6^\circ) + 0.4 \quad (6)$$

$$^3J_{\text{Munzarová}}(\text{C8} - \text{H1}') = 4.2 \cos^2(\chi' - 68.9^\circ) - 0.5 \cos(\chi' - 68.9^\circ) + 0.3 \quad (7)$$

We note that the φ phase shift of $+60^\circ$ ($+68.9^\circ$ in eq 7) used consistently in Karplus curves for both the $^3J(\text{C4}-\text{H1}')$ and $^3J(\text{C8}-\text{H1}')$ couplings (as was suggested first by Ippel¹⁰ and Wijmenga¹) instead of the -120° shift for the later J coupling corresponds to the negative sign of the B coefficient (eqs 3, 5, and 7). The amplitudes and phase shifts (Tables 1 and 2) were fitted for the calculated data (Supporting Information) with the Levenberg–Marquardt procedure combined with the simplex iteration method using the Microcal(TM) Origin 6.0 program. The interval calculated for the χ torsion ($0-360^\circ$) was expanded symmetrically by one 360° period before and after this interval prior to the fitting. The overall quality of the fitted parameters shown in Tables 1 and 2 was very good ($R^2 = 0.9881-0.9980$, $\chi^2 = 0.0330-0.0073$). We note that the $^3J(\text{C4}-\text{H1}')$ coupling was assigned to the χ torsion, while the $^3J(\text{C8}-\text{H1}')$ coupling was fitted as dependent on the χ' torsion. The χ and χ' torsion angles are far from being the same for different sugar-to-base orientations, and therefore, the $^3J(\text{C8}-\text{H1}')$ coupling has to be assigned to the χ' torsion, which directly corresponds to the spin-spin coupling pathway (Figures 1 and 4).

TABLE 1: The Parameters A , B , and C in Hz and the Phase φ in Degrees of Karplus Equations for the $^3J(\text{C4}-\text{H1}')$ Coupling Assigned to the $\text{O4}'-\text{C1}'-\text{N9}-\text{C4}$ Dihedral Angle (the glycosidic torsion χ)

nucleoside	sugar conformation	solvent	A	B	C	φ
B-DNA	C2'-endo		3.68	1.87	0.49	70.44
B-DNA	C3'-endo		3.66	1.84	0.41	71.11
B-DNA	C2'-endo	water	3.82	1.99	0.60	69.71
B-DNA	C3'-endo	water	3.82	1.95	0.53	70.56
Z-DNA	C2'-endo		3.82	1.85	0.42	70.62
Z-DNA	C3'-endo		3.61	1.78	0.45	68.26
A-RNA	C2'-endo		3.50	1.73	0.51	70.09
A-RNA	C3'-endo		3.86	1.83	0.41	71.68
B-DNA ^a	C2'-endo		3.6	1.8	0.4	68.9

^a Reference 16.

TABLE 2: The parameters A , B , and C in Hz and the Phase φ in Degrees of Karplus Equations for the $^3J(\text{C8}-\text{H1}')$ Coupling Assigned to the $\text{H1}'-\text{C1}'-\text{N9}-\text{C8} + 60^\circ$ Dihedral Angle

nucleoside	sugar conformation	solvent	A	B	C	φ
B-DNA	C2'-endo		4.44	-0.60	0.29	68.89
B-DNA	C3'-endo		4.75	-1.09	0.18	69.90
B-DNA	C2'-endo	water	4.57	-0.77	0.39	68.92
B-DNA	C3'-endo	water	4.96	-1.30	0.29	69.31
Z-DNA	C2'-endo		4.34	-0.71	0.31	67.58
Z-DNA	C3'-endo		4.57	-1.07	0.37	67.40
A-RNA	C2'-endo		4.34	-0.66	0.27	69.71
A-RNA	C3'-endo		5.12	-1.22	0.14	73.06
B-DNA ^a			4.2	-0.5	0.3	68.9

^a Reference 16.

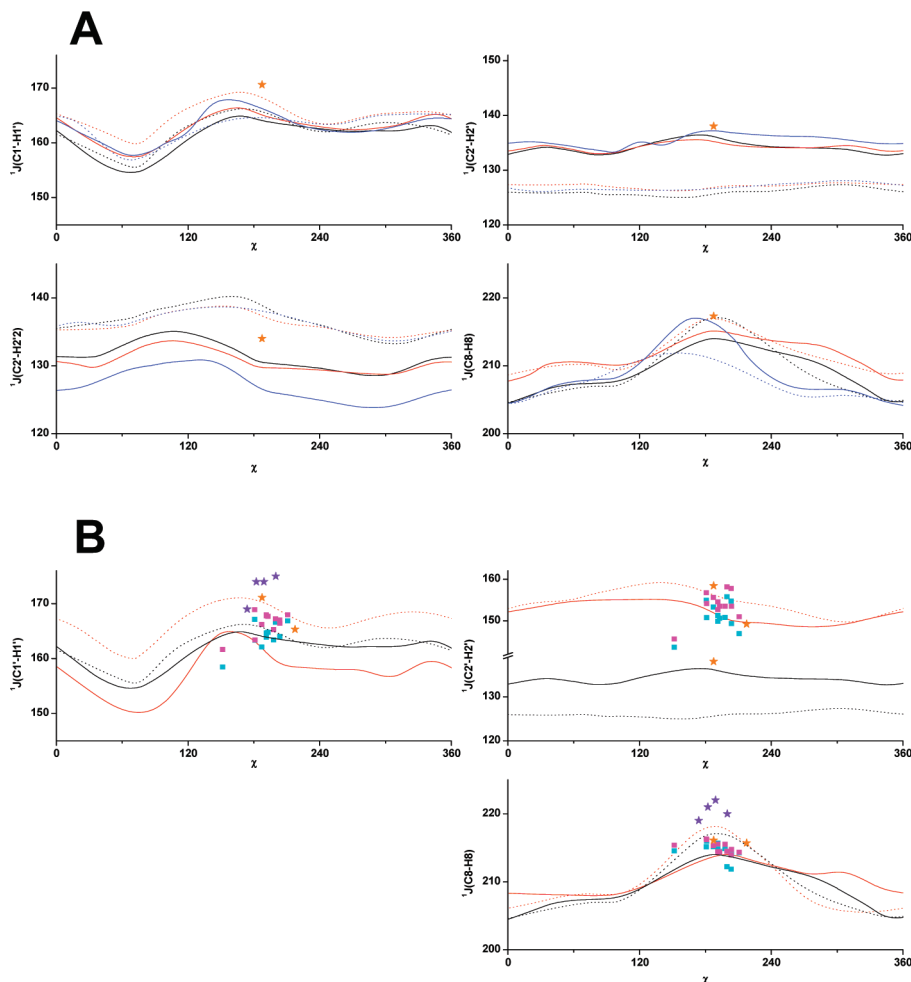


Figure 5. The calculated dependence for the $^1J(\text{C1}'\text{--H1}')$, $^1J(\text{C2}'\text{--H2}')$, $^1J(\text{C1}'\text{--H2'2})$, and $^1J(\text{C8}\text{--H8})$ coupling constants in Hz plotted as function of the χ (the glycosidic torsion) angle in degrees. The C2'- and C3'-endo sugar pucker conformations correspond to the solid and dotted lines, respectively. (A) The dG nucleosides; B-DNA without solvent (black), B-DNA including PCM solvent (red), and Z-DNA without solvent (blue). (B) The rG nucleosides; B-DNA without solvent added for comparison (black), A-RNA without solvent (red), and WC/SE base pairs without and including water solvent (blue and magenta squares, respectively). The experimental 3J couplings in DNA (orange asterisk¹⁰) and RNA (orange¹⁰ and violet⁴² asterisk).

Different sugar pucker of the dG and rG nucleosides had an effect on the fitted amplitudes in the Karplus equations for the $^3J(\text{C8}\text{--H1}')$ coupling. The absolute amplitudes A and B fitted in the C2'-endo nucleosides were consistently smaller compared to those calculated for the C3'-endo nucleosides (Table 1). The amplitudes obtained for the $^3J(\text{C4}\text{--H1}')$ coupling were more independent of sugar pucker, except for the dG nucleoside with the Z-DNA backbone (Table 2). The Karplus equations fitted here for the two sugar pucker reflect the trend calculated by Munzarová.¹⁶

Application of the implicit water solvent led to the increase of the $^3J(\text{C4}\text{--H1}')$ and $^3J(\text{C8}\text{--H1}')$ couplings relative to the gas-phase values by as much as 0.16 and 1 Hz, respectively. This increase was reflected by the increase of absolute magnitudes of the A and B amplitudes (Tables 1 and 2).

The fitted φ phase shift ranged from 67.4 to 73.1°, which is in agreement with the values calculated by Munzarová¹⁶ (Tables 1 and 2). Also, the fitted amplitudes for the two 3J couplings are close to those which were obtained by Munzarová.¹⁶ This together indicated that (a) the φ phase shift should be included as a “free” parameter into the fitting procedure for the $^3J/\chi$ Karplus equations and (b) the obtained Karplus equations are sufficiently robust with respect to the geometry and solvent effects modeled in this work.

3.3. The 1J Couplings. The $^1J(\text{C1}'\text{--H1}')$ and $^1J(\text{C8}\text{--H8})$ couplings should depend dominantly on the glycosidic torsion, while the $^1J(\text{C2}'\text{--H2}')$ and $^1J(\text{C2}'\text{--H2'2})$ ones should depend dominantly on the sugar pucker.^{1,10,39}

The $^1J(\text{C1}'\text{--H1}')$ couplings calculated in the dG and rG nucleosides ranged from 154 to 170 Hz. The smaller magnitudes were calculated in the *syn*-nucleosides, while the maximal magnitudes of the J coupling were calculated near the *anti* region for $\chi \approx 180^\circ$ (Figure 5). The magnitudes of the $^1J(\text{C1}'\text{--H1}')$ coupling measured in the *anti*-guanosines by Kline⁸ or Varani and Tinoco⁴⁰ and also in other nucleosides^{8,10,39,41} were similar to the calculated range. Specificity of the NA base should not therefore significantly affect the $^1J(\text{C1}'\text{--H1}')$ coupling magnitude.

Application of the implicit water solvent led to the increase of the calculated $^1J(\text{C1}'\text{--H1}')$ coupling by as much as 4.4 Hz in the *syn* region (Figure 5). In the *anti* region, near $\chi \approx 240^\circ$, the effect of implicit solvent was damped down. This happened probably due to the close spatial proximity of the H1' and N3 atoms, which may hinder effective polarization of the C1'–H1' bond by the implicit solvent cavity. Magnitudes of the $^1J(\text{C1}'\text{--H1}')$ couplings calculated in the C3'-endo nucleosides were overall larger roughly by 2 Hz in comparison with those calculated in the C2'-endo nucleosides. The larger effect of sugar pucker was calculated in the *syn* region of the rG nucleosides

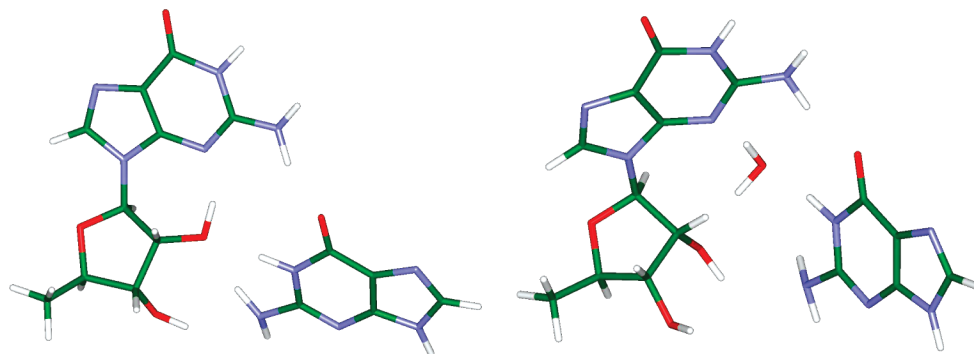


Figure 6. Molecular complexes of the *cis*-Watson-Crick/sugar edge G/rG base and its water-mediated substate (right).²⁵

compared to the dG ones (Figure 4A, B). The $^1J(\text{C1}'\text{--H1}')$ couplings calculated in the rG nucleoside were smaller at most by 5 Hz than those calculated in the dG one, except for the region with $\chi \approx 180^\circ$ where the magnitudes calculated in both nucleosides were almost the same (Figure 5B).

Except for the parametrization by Munzarová¹⁶ for the $^1J(\text{C1}'\text{--H1}')$ coupling,¹⁶ no other parametrizations for the Karplus equations for any of the four 1J couplings were reported. Although the overall trend calculated for the $^1J(\text{C1}'\text{--H1}')$ coupling is very similar to the trend obtained by Munzarová, the magnitudes calculated by Munzarová are systematically smaller approximately by 10 Hz. Since our calculations of the $^1J(\text{C1}'\text{--H1}')$ coupling correspond well with the experiment,^{10,36,42} the underestimated trend calculated previously by Munzarová was most probably caused by some systematic error which is difficult for us to clarify since they used a different calculation method and software.

The calculated $^1J(\text{C8--H8})$ couplings ranged from 204 to 218 Hz, which is in qualitative agreement with a few experiments available for this 1J couplings in the literature reporting the range from 215.7 to 222 Hz.^{10,11} The calculated dependence on the χ torsion had one maximum near 180° and one minimum near $0/360^\circ$ (Figure 5). The two extremes of the calculated dependence actually correspond to the planar arrangement of the H8-C8-N9-C1'-O4' covalently linked atoms. The C2'/C3'-endo sugar pucker effect on the calculated $^1J(\text{C8--H8})$ couplings was opposite in the decaying part of the dependence and near the maximum where the curves calculated in the C3'-endo nucleosides actually dominated (at most by 3 Hz) over those calculated in the C2'-endo ones (Figure 5A). The $^1J(\text{C8--H8})$ couplings calculated for the $\chi \approx 0^\circ$ in the rG nucleoside were smaller than those in the dG nucleoside, but near $\chi \approx 180^\circ$, both calculated dependencies almost coincided (Figure 5B). The $^1J(\text{C8--H8})$ coupling measured in the dG and rG nucleoside of the same molecule, the circular r< pGp(dG)> dinucleoside, was 216.1 and 217.3 Hz, respectively.¹⁰ Both nucleosides in the r< pGp(dG)> molecule had the same torsion angle of $\chi = 187.5^\circ$ (SIWWOK in the Cambridge Database^{43,44}). Practically the same magnitudes of the $^1J(\text{C8--H8})$ coupling measured in the r< pGp(dG)> molecule with a well-resolved value of the χ torsion correlate nicely with our calculations.

Other values of the $^1J(\text{C8--H8})$ coupling measured in guanosines were 215.7 Hz in cyclic trinucleoside,¹⁰ 216.0 Hz in guanosine monophosphate,¹⁰ and 214.85 Hz in the RNA oligomer from the helix-35 of *E. coli* 23S rRNA.¹¹

The $^1J(\text{C8--H8})$ couplings calculated with the inclusion of implicit water solvent were overall larger than those obtained in the gas-phase calculations (at most by 4 Hz) and ranged from 207 to 218 Hz. This is in agreement with our previous study of the $^1J(\text{C8--H8})$ coupling in the DNA hairpin molecule.¹⁷

Modifications of the β and γ torsion angles (Figure 1) in the dG nucleosides, which reflects their orientation in the B- or Z-DNA backbone, had inconsiderable impact on the curvature of the calculated dependences (Figure 5A).

The calculated $^1J(\text{C2}'\text{--H2}')$ and $^1J(\text{C2}'\text{--H2}''')$ couplings strongly depended on the sugar type and conformation (Figure 5).

The $^1J(\text{C2}'\text{--H2}')$ couplings calculated in the dG nucleoside (the gas-phase calculation) ranged from 132.4 to 137.2 Hz and from 124.7 to 128.5 Hz for the C2'- and C3'-endo sugar puckers, respectively. Similar separation of the dependencies with different sugar pucker was calculated also for the $^1J(\text{C2}'\text{--H2}''')$ couplings ranging from 123.7 to 135.2 Hz (C2'-endo), and from 133.0 to 140.7 Hz (C3'-endo). The calculated trends agreed with the $^1J(\text{C2}'\text{--H2})$ and $^1J(\text{C2}'\text{--H2}''')$ couplings measured by Ippel.¹⁰ The effect of the implicit solvent on the calculated magnitudes of both 1J couplings was negligible in comparison with the effects of sugar pucker and backbone torsion variation.

A different kind of sugar in the dG and rG nucleosides had a large impact on the calculated $^1J(\text{C2}'\text{--H2})$ and $^1J(\text{C2}'\text{--H2}''')$ couplings (Figure 5B). For the $^1J(\text{C2}'\text{--H2}')$ coupling in the dG and rG nucleoside, the calculated mean values differed by 20 and 29 Hz for the C2'- and C3'-endo sugar puckers, respectively. This trend calculated for the $^1J(\text{C2}'\text{--H2}')$ coupling agrees with the 138 ± 2 and 158.4 ± 0.5 Hz couplings measured in the r< pGp(dG)> molecule for the dG and rG nucleoside,¹⁰ respectively. The $^1J(\text{C2}'\text{--H2}')$ coupling of 138 Hz and the $^1J(\text{C2}'\text{--H2}''')$ one of 134 Hz, which were measured in the r< pGp(dG)> molecule, would therefore, according to our calculations, indicate the C2'-endo sugar pucker of the dG nucleoside rather than C3'-endo one (Figure 5).

3.4. Impact of Base Pairing and Explicit Water Molecules on the Magnitude of the J Constants. Base pairing belongs to the essential interactions which stabilize the structure of the polynucleotides in DNA and RNA molecules. Modeling the effect of various base pairing interactions on J coupling magnitudes would therefore confront the trends obtained in bare nucleosides with those in realistic structural patterns occurring in NAs.

As a structural models for such a theoretical NMR study, we adopted the RNA Watson-Crick/sugar edge (WC/SE)²² class of base pairs (Figures 3 and 6), which was also intensively studied previously.^{26,45} We selected actually the 12 WC/SE base pairs in which the rG nucleoside interacts via its SE with all other nucleobases. This class of extended RNA base pairing interactions is highly relevant for the purpose since (a) the noncovalent interactions in the WC/SE base pairs occur in the proximity of the glycosidic bond, in contrast to the canonical WC/WC interactions, and (b) WC/SE base pairs (and other types of SE base pairs) are common in ribosome and other functional RNAs, as evidenced by the numerous X-ray studies. They are,

in fact, essential for the formation of many functional RNA building blocks and motifs. Thus, characterizing their NMR signatures by calculations can provide an important link between theory and experiment, which can improve the NMR structural analysis of RNAs.

Variation of the glycosidic torsion angle in the WC/SE base pairs (from 150 to 210°) covers the low anti region. The calculated variation of the $^3J(\text{C8-H1}')$, $^3J(\text{C4-H1}')$, $^1J(\text{C1'-H1}')$, $^1J(\text{C2'-H2}')$, and $^1J(\text{C8-H8})$ couplings in the WC/SE base pairs was 0.2–3.1, 0.3–1.2, 158.5–169.0, 143.7–158.2, and 211.9–216.3 Hz, respectively. When the WC/SE base pairs were embedded into the implicit water solvent, the absolute magnitudes of the coupling increased very little (Figures 4B and 5B).

The $^3J(\text{C4-H1}')$ couplings calculated in the WC/SE base pairs fit perfectly the Karplus curves (Figure 4B). The $^3J(\text{C8-H1}')$ couplings in the WC/SE base pairs probably had more disperse character, but they still fit the Karplus curve very well. Interestingly, the carbon C8 seemed to be more perturbed by the WC/SE base pair interactions than the C4 one, although its spatial proximity to the “perturbing” base in a pair was larger.

The trends calculated for the three 1J couplings in the rG nucleoside and in the WC/SE base pairs were not qualitatively equivalent. The large impact of the base pairing was calculated especially for the $^1J(\text{C1'-H1}')$ and $^1J(\text{C2'-H2}')$ couplings, where the atoms coupled in the spin–spin interaction and the sugar atoms exposed to the sugar edge of the WC/SE base pairs were actually the same (Figures 4B and 5B). If any structural information concerning the glycosidic torsion in the rG nucleosides can be acquired from the three 1J couplings, it could be easily lost due to the large effect and individual character of the noncovalent interactions in the sugar edge.

Some WC/SE base pairs (and other base pairs involving the sugar edges) had water-mediated contacts between the ribose and base (Figure 6). The balance between direct and water-mediated substates of the base pairs may be important for RNA structural dynamics.⁴⁶ The effect of such an explicit water molecule on the J couplings can be estimated from the calculations in the *cis*-G/rG base pair (Figure 6). The glycosidic torsion calculated in the *cis*-G/rG complex without and with water molecule was 151 and 181°, respectively. The calculated absolute shift of J couplings due to the explicit water in the *cis*-G•rG complex was smaller than 0.6 Hz in the case of the two 3J couplings and the $^1J(\text{C8-H8})$ coupling, while for the $^1J(\text{C1'-H1}')$ and $^1J(\text{C2'-H2}')$ couplings, it was 8.7 and 11.2 Hz, respectively. The trends calculated for the $^1J(\text{C1'-H1}')$ and $^1J(\text{C2'-H2}')$ couplings may therefore require specific calibration, which would be dependent on the topology and mobility of the water solvent molecules that can penetrate into the NA base pairs.

4. Conclusion

We have carried out a comprehensive computational study of the dependence of the $^3J(\text{C8-H1}')$, $^3J(\text{C4-H1}')$, $^1J(\text{C1'-H1}')$, $^1J(\text{C8-H8})$, $^1J(\text{C2'-H2}')$, and $^1J(\text{C2'-H2}')$ NMR spin–spin coupling constants on the geometry in the dG (deoxyguanosine) and rG (riboguanosine) molecules.

Calculations of the six J couplings were focused on their dependence on the following geometry parameters: the glycosidic torsion angle χ , the sugar pucker C2'- and C3'-endo, the β and γ torsion angles of the hydroxymethyl group at the C4' carbon of sugar used for modeling the effect of different NA backbone geometry (B- and Z-DNA), and the type of sugar (dG, rG). The effect of the base pairing on J couplings was calculated for the rG nucleosides in the WC/SE RNA base pairs.

The $^3J(\text{C8-H1}')$ and $^3J(\text{C4-H1}')$ couplings depended dominantly on the χ torsion. Sugar pucker of both the dG and rG nucleosides had systematic impact on the calculated χ dependencies of the $^3J(\text{C8-H1}')$ coupling (the larger magnitudes were obtained in the C3'-endo nucleosides), while the dependence of $^3J(\text{C4-H1}')$ coupling on sugar pucker was practically negligible (cf. also Munzarová et al.¹⁶). New Karplus equations respecting sugar pucker in the dG and rG nucleosides taking into account the effect of water solvent were fitted for the two 3J couplings. The calculated impact of the backbone geometry on the two 3J couplings was practically negligible, as well as the effect of base pairing in the WC/SE RNA base pairs. This fact better validates use of these Karplus equations in NMR structural studies.

The calculated $^1J(\text{C2'-H2}')$ and $^1J(\text{C2'-H2}')$ couplings depended dominantly on the type of sugar in the rG and dG nucleosides and also on its sugar pucker. The $^1J(\text{C1'-H1}')$ and $^1J(\text{C8-H8})$ couplings were more dependent on the glycosidic torsion angle, and qualitatively different dependencies were calculated for the dG and rG nucleosides. The structural information concerning the χ torsion acquired from the three 1J couplings can be strongly affected by the specificity of the noncovalent interactions in the sugar edge of the rG nucleosides. Calculated dependences of the 1J couplings thus brought only additional information regarding the possible determination of the glycosidic torsion, which can be nevertheless used as support in the context of better-suited $^3J(\text{C4/C8-H1}')$ couplings. The three 1J couplings, however, still represent another, although less significant, NMR structural constraint for the NMR structural studies focused on the glycosidic torsion. For example, the $^1J(\text{C8-H8})$ coupling changes roughly by 10 Hz when the guanine base adopts orientation with respect to sugar, corresponding to $\chi = 0$ and 180°. The dependences calculated for the six J couplings on the χ torsion are in good overall agreement with available experimental data found in the literature.

Acknowledgment. This work was supported by the Academy of Sciences of the Czech Republic (Grant IAA400550701), by the Grant Agency of Charles University (Grant GAUK nr. 58708), by The Netherlands Organization for Scientific Research (NWO-CW), and by the National Research School Combination - Catalysis (NRSC-C). V.S. was supported by a Human Frontier Science Program (HFSP) Young Investigator's Grant. J.S. was also supported by the Academy of Sciences of the Czech Republic, Grants AV0Z50040507 and AV0Z50040702.

Supporting Information Available: All calculated J couplings and corresponding geometry parameters. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. NMR Spectrosc.* **1998**, 32, 287.
- (2) Alkorta, I.; Elguero, J. *Int. J. Mol. Sci.* **2003**, 4, 64.
- (3) Sychrovský, V.; Vokáčová, Z.; Šponer, J.; Špačková, N.; Schneider, B. *J. Phys. Chem. B* **2006**, 110, 22894.
- (4) Correll, C. C.; Wool, I. G.; Munishkin, A. *J. Mol. Biol.* **1999**, 292, 275.
- (5) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, 289, 905.
- (6) Wahl, M. C.; Sundaralingam, M. *Biopolymers* **1997**, 44, 45.
- (7) Ng, H. L.; Dickerson, R. E. *Nucleic Acids Res.* **2002**, 30, 4061.
- (8) Kline, P. C.; Serianni, A. S. *J. Am. Chem. Soc.* **1990**, 112, 7373.
- (9) Davies, D. B.; Rajani, P.; Maccoss, M.; Danyluk, S. S. *Magn. Reson. Chem.* **1985**, 23, 72.

- (10) Ippel, J. H.; Wijmenga, S. S.; de Jong, R.; Heus, H. A.; Hilbers, C. W.; de Vroom, E.; van der Marel, G. A.; van Boom, J. H. *Magn. Reson. Chem.* **1996**, *34*, S156.
- (11) Boisbouvier, J.; Bryce, D. L.; O'Neil-Cabello, E.; Nikonowicz, E. P.; Bax, A. *J. Biomol. NMR* **2004**, *30*, 287.
- (12) Hines, J. V.; Landry, S. M.; Varani, G.; Tinoco, I. *J. Am. Chem. Soc.* **1994**, *116*, 5823.
- (13) Marino, J. P.; Schwalbe, H.; Griesinger, C. *Acc. Chem. Res.* **1999**, *32*, 614.
- (14) Zimmer, D. P.; Marino, J. P.; Griesinger, C. *Magn. Reson. Chem.* **1996**, *34*, S177.
- (15) Munzarová, M. L.; Sklenář, V. *J. Am. Chem. Soc.* **2002**, *124*, 10666.
- (16) Munzarová, M. L.; Sklenář, V. *J. Am. Chem. Soc.* **2003**, *125*, 3649.
- (17) Sychrovský, V.; Schneider, B.; Hobza, P.; Židek, L.; Sklenář, V. *Phys. Chem. Chem. Phys.* **2003**, *5*, 734.
- (18) Trantírek, L.; Štefl, R.; Masse, J. E.; Feigon, J.; Sklenář, V. *J. Biomol. NMR* **2002**, *23*, 1.
- (19) Foloppe, N.; Nilsson, L.; MacKerell, Jr., A. D., Jr. *Biopolymers* **2002**, *61*, 61.
- (20) Sychrovský, V.; Muller, N.; Schneider, B.; Smrecki, V.; Špirko, V.; Šponer, J.; Trantírek, L. *J. Am. Chem. Soc.* **2005**, *127*, 14663.
- (21) Bouř, P.; Buděšínský, M.; Špirko, V.; Kapitán, J.; Šebestík, J.; Sychrovský, V. *J. Am. Chem. Soc.* **2005**, *127*, 17079.
- (22) Leontis, N. B.; Stombaugh, J.; Westhof, E. *Nucleic Acids Res.* **2002**, *30*, 3497.
- (23) Leontis, N. B.; Westhof, E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 300.
- (24) Vallurupalli, P.; Moore, P. B. *J. Mol. Biol.* **2003**, *325*, 843.
- (25) Šponer, J. E.; Špačková, N.; Kulhánek, P.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. A* **2005**, *109*, 2292.
- (26) Šponer, J. E.; Špačková, N.; Leszczynski, J.; Šponer, J. *J. Phys. Chem. B* **2005**, *109*, 11399.
- (27) Kaupp, M.; Buhl, M.; Malkin, V. *Calculation of NMR and EPR Parameters—Theory and Applications*; Wiley-VCH Verlag: Weinheim, Germany, 2004.
- (28) Helgaker, T.; Jaszunski, M.; Ruud, K. *Chem. Rev.* **1999**, *99*, 293.
- (29) Sychrovský, V.; Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2000**, *113*, 3530.
- (30) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.
- (31) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR - Basis Principles and Progress*; Springer: Heidelberg, Germany, 1990.
- (32) Cammi, R.; Mennucci, B.; Tomasi, J. *J. Phys. Chem. A* **2000**, *104*, 5631.
- (33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (34) Zhu, G.; Live, D.; Bax, A. *J. Am. Chem. Soc.* **1994**, *116*, 8370.
- (35) Schmieder, P.; Ippel, J. H.; Vandenberg, H.; Vandermarel, G. A.; Vanboom, J. H.; Altona, C.; Kessler, H. *Nucleic Acid Res.* **1992**, *20*, 4747.
- (36) Schwalbe, H.; Marino, J. P.; King, G. C.; Wechselberger, R.; Bermel, W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 631.
- (37) Karplus, M. *J. Chem. Phys.* **1959**, *30*, 11.
- (38) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870.
- (39) Davies, D. B.; Maccoss, M.; Danyluk, S. S. *J. Chem. Soc., Chem. Commun.* **1984**, 536.
- (40) Varani, G.; Cheong, C. J.; Tinoco, I. *Biochemistry* **1991**, *30*, 3280.
- (41) Bandyopadhyay, T.; Wu, J.; Stripe, W. A.; Carmichael, I.; Serianni, A. S. *J. Am. Chem. Soc.* **1997**, *119*, 1737.
- (42) Varani, G.; Tinoco, I. *J. Am. Chem. Soc.* **1991**, *113*, 9349.
- (43) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr.* **1979**, *B35*, 2331.
- (44) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187.
- (45) Vokáčová, Z.; Šponer, J.; Šponer, J. E.; Sychrovský, V. *J. Phys. Chem. B* **2007**, *111*, 10813.
- (46) Rázga, F.; Koča, J.; Šponer, J.; Leontis, N. B. *Biophys. J.* **2005**, *88*, 3466.

JP902473V

Computational Study on Spectral Properties of the Selected Pigments from Various Photosystems: Structure–Transition Energy Relationship

Zuzana Vokáčová and Jaroslav V. Burda*

Department of Chemical Physics and Optics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 3, 121 16 Prague 2, Czech Republic

Received: February 28, 2007; In Final Form: April 21, 2007

In this study, the most important kinds of pigments (chlorophylls, bacteriochlorophylls, phycobilins, and carotenoids) from various photosystems were explored. For the most stable conformations, electronic transitions were determined at the TDDFT/6-31+G(d) level with the B3PW91 functional and compared to measured spectra. The group of carotenoids was also investigated at the TDA/TDDFT level with the BLYP functional. The energies of Q_y transitions are systematically blue-shifted by about 50–100 nm in the case of (bacterio)chlorophyll and pheophytin molecules. Nevertheless, the correct relative order of the Q lines among various chlorophyll types was obtained through comparison with experimental results. Much better agreement was obtained for the Soret band, for which the differences between calculated and measured transitions were at most 35 nm. In the case of phycobilins, the first transition line was estimated to be at lower frequencies (around 500 nm) with a very similar blue shift of about 100 nm from experimental values. The influence of anchoring cysteine side chain(s) was found to be marginal. A dominant effect of the linear polyene chain on the determined spectral lines was found in the case of carotenoids. Nevertheless, the impact of β -cycles and epoxy and keto groups is clearly visible as well. The high intensity of the first allowed transition matches different characters of the HOMO and LUMO. In the case of fucoxanthin, the TDA method also predicts the B_u^- state to lie below the $1B_u^+$ state. Because the shift of electron transitions is approximately proportional to the size of the π -conjugated system, the shift of the calculated transitions compared to experimental values is practically constant for the same excitations of (bacterio)chlorophyll and phycobilin molecules. However, this is not true for carotenoids, for which both the transition energy and the shift of the transition vary with the number of conjugated double bonds.

Introduction

Photosynthetic pigments represent the lifeblood of all photoautotrophic organisms. There are basically three classes of photosynthetic pigments: (bacterio)chlorophylls, carotenoids, and the smaller but still important phycobilins.¹ These pigments have been studied for many years, and several investigators have been recognized with Nobel Prizes in the distant and even recent past for elucidating the importance of the subject. Despite the great effort devoted to this subject, there is still much room for further exploration. We decided to examine the electron transition spectra of individual structures and the relationship between structure and transition energy in the above-mentioned pigment classes.

Experimental electronic spectra of these compounds can be found in many papers, monographs,² and spectral encyclopedias.³ Because this work has a computational character, we concentrate here mainly on the theoretical studies dealing with spectra predictions (or estimations) performed on the chosen pigments and their models, using measured spectra for comparison and approximate error estimation.

For (bacterio)chlorophylls, the very first calculations were performed by Gouterman and co-workers^{4–6} on the porphyrin ring, introducing the four-orbital model. Following these initial efforts, many other researchers have performed calculations at various levels using different models frequently based on free-

base porphyrin; see, for instance, refs 7–24. These calculations are usually used for method-testing purposes, and up to now, they have usually been too demanding for larger pigment molecules or their aggregates. The electronic spectra of chlorophyll *a* have been explored in many studies, including those of Hasegawa et al.,²⁵ who used the relatively high-accuracy SAC-CI (symmetry-adapted cluster-configuration interaction) method, and Parusel and Grime,²⁶ who compared chlorophyll *a* and pheophytin. The orientation of two porphyrin rings in a dimeric arrangement was investigated and compared with experimental data (Raman spectroscopy) by Jeong et al.²⁷

Spectroscopic properties of some bacteriochlorophylls (*a*, *b*, *c*1–3, and *d*) were studied using a semiempirical approach by Linnato et al.^{28,29} Later, they extended the explored systems to self-organized aggregates of bacteriochlorophylls³⁰ and the LH II antenna structure.³¹ Cory et al.³² determined semiempirically electronic excitations in larger aggregates of bacteriochlorophylls. Accurate TDDFT (time-dependent density functional theory) calculations of chlorophyll *a* and bacteriochlorophyll *b* spectra have been published by Sundholm.^{33,34}

In the case of (bacterio)chlorophyll molecules, for a comparison with experimental data, we used measurements of absorption spectra,^{35,36} circular dichroism,³⁷ and fluorescence and absorption spectra.^{38,39}

Phycobilins belong to the second group of pigments that are present in photosynthetic complexes. Phycobilins are linear open-chain tetrapyrrole systems anchored in proteins. An

* Corresponding author. E-mail: burda@karlov.mff.cuni.cz.

interesting recent study⁴⁰ addressed the spectral characterization of phycoerythrin in combination with quantum chemical calculations. Another study investigated regulation of the absorbed excitation energy by chlorophyll and phycobilin. Their models are based on X-ray structures of PS I, PS II, and allophycocyanin.⁴¹

Carotenoids are the last family of photosynthetic pigments that we examine in this study. Molecular structures of selected carotenoids were previously calculated semiempirically (AM1) by Hashimoto et al.⁴² The obtained structures were compared with geometries from X-ray crystallography. Similarly, in the work of Wang et al.,⁴³ structural characteristics and (stacking) stabilization energies were explored using the MP2 approach. The role of beta-carotene in the quenching of singlet oxygen was examined in a study⁴⁴ using the DFT method. Great effort has also been devoted to calculations of electronic spectra. In the study of He et al.,⁴⁵ the complete active space SCF (CASSCF) technique was employed on the low-lying excited states of the carotenoid rhodopin glucoside (RG). The results showed that the lowest triplet state energy of the RG is below 0.78 eV. The authors also found a significant blue shift of the energy of the second excited singlet state (S2) in this carotenoid under the influence of bacteriochlorophylls. In another work,⁴⁶ theoretical and experimental approaches were combined to demonstrate some correlations between AM1 calculations and NMR spectroscopy. A large number of studies on carotenoid spectra have been published by the Fleming group,^{47–61} who used both experimental and theoretical tools. Their quantum chemical calculations were done in collaboration with the group of Head-Gordon.^{62–64} In one recent studies,⁶⁰ TDDFT and TDA/TDDFT (Tamm–Dancoff approximation)⁶⁵ calculations were used for the exploration of the energy transfer from the S1 state of peridinin to chlorophyll in photosystem (PS) I.

Although the TDA cannot generally be considered superior to the full TDDFT method, in carotenoids, it performs better because of the fact that full TDDFT overestimates the interactions of close-lying states, as pointed in a study by Hirata and Head-Gordon.⁶⁵

Failures of the TDDFT method applied to dimers of pigments from photosystems and their models have been published, e.g., by Dreuw et al.⁶⁶ The problem results from unphysical charge transfer, which generates low-lying “ghost states”. A remedy was recently suggested by Cai et al.⁶⁷

In the present study, a systematic determination of TDDFT electronic spectra was performed for the most common (bacterio)chlorophylls, phycobilins, and carotenoids. The main goal of this study was a comparison of pure electron transition energies of isolated (optimal gas-phase) structures without any influence of neighboring peptide molecules, other pigments, or solvation effects using the same method and basis set. In this way, all undesired shifts can be suppressed, and genuine characteristics of individual structures can be elucidated.

Computational Details

First, several conformers of each pigment were chosen to find a global minimum. A structural database [Protein Data Bank⁶⁸ (pdb)] was used to obtain appropriate initial structures.

This selection was important in the case of the various conformers of chlorophyll (types *a*, *b*, *c*1–3, *d*) and bacteriochlorophyll (types *a*, *b*, *c*, *d*, *e*, *g*), whose structures are presented in Figures 1 and 2, respectively. Usually, several arrangements were possible, and it was not completely clear which could be excluded based on an exploration of the structural database because the described conformations usually depend strongly

on the corresponding biological environment. Also, several structures of a given (bacterio)chlorophyll type were found to have similar total energies. One of the rare exceptions, for which several high-resolution X-ray structures are available, is the molecule of chlorophyll *a*. Our model structures were adopted from 1DOP pdb structure. In all investigated (bacterio)chlorophyll conformers, the phytol chain was excluded from consideration. Instead, the methyl ester anchoring group was used in the models. This can be assumed to be a reasonable approximation, as it follows from our preliminary calculations as well as from other studies (e.g., ref 69). For the other chlorophyll types, the original chlorophyll *a* ligands were replaced by appropriate substituents to create the required models. In the determination of pheophytin *a*, the structure of chlorophyll *a* was taken as a starting point, with the Mg²⁺ cation substituted by two protons.

In the case of the phycobilin group, two representatives were considered, as shown in Figure 3. The initial structures were obtained from the Protein Data Bank.⁶⁸ Because phycobilins can adopt many spatial conformations, the crystallographic structure of phycocyanobilin was chosen as a starting point to preserve the biologically relevant arrangement of the pigment. The models of phycoerythrobilin originate from the same starting structure with replacement of the different substituents and reduction of the double bond between the C and D pyrrole rings. In the case of phycoerythrobilin, two forms were considered: anchored through a single covalent bond to the sulfur atom of cysteine side chain and/or with a pair of covalent bonds as depicted in Figure 3.

In the case of carotenoids, nine pigments were chosen; they are displayed in Figure 4. Selection of these pigments allows for clarification of the influence of different structural elements on spectral characteristics. The all-trans conformation was used for linear carbon chains. The other substituents (e.g., xanthophyll and β -cycles) were constructed according to the basic “hybridization” rules, and the lowest-energy conformers were chosen for subsequent analysis.

Because all of the conformers represent quite an extended set of large molecules, several subsequent geometry optimizations were performed, making the selected set of conformers gradually smaller. The lowest-energy structures were finally reoptimized at the B3PW91/6-31G(d) level of theory. For these molecules, absorption spectra were determined with time-dependent density functional theory (TDDFT) using the same functional and the 6-31+G(d) basis set. The B3PW91 functional and basis set were chosen on the basis of our previous calculations on free-based porphine systems.^{70,71} The diffuse functions were found to be of key importance for the determination of electron transitions, as described in many other studies (e.g., refs 72 and 73).

All calculations were performed using the Gaussian 03 program package.

Pigment Isomers/Conformers. Many relevant structures were considered in the first optimization step (HF/3-21G). In the next step, reduced sets of conformers (usually fewer than five) were considered, and the lowest-energy structures were reoptimized at the B3PW91/6-31G(d) level.

In the case of chlorophyll *b*, a pair of conformers with *cis* and *trans* orientations of the C7=O carbonyl bond in the formaldehyde group with respect to the C8 site (for the atom numbering, see Figure 5) exhibits a small energy difference (about 1 kcal/mol). Global minimum of chlorophyll *b* represents conformer with the oxygen from the –COH group in a *trans* arrangement with respect to the C8 atom. In this orientation,

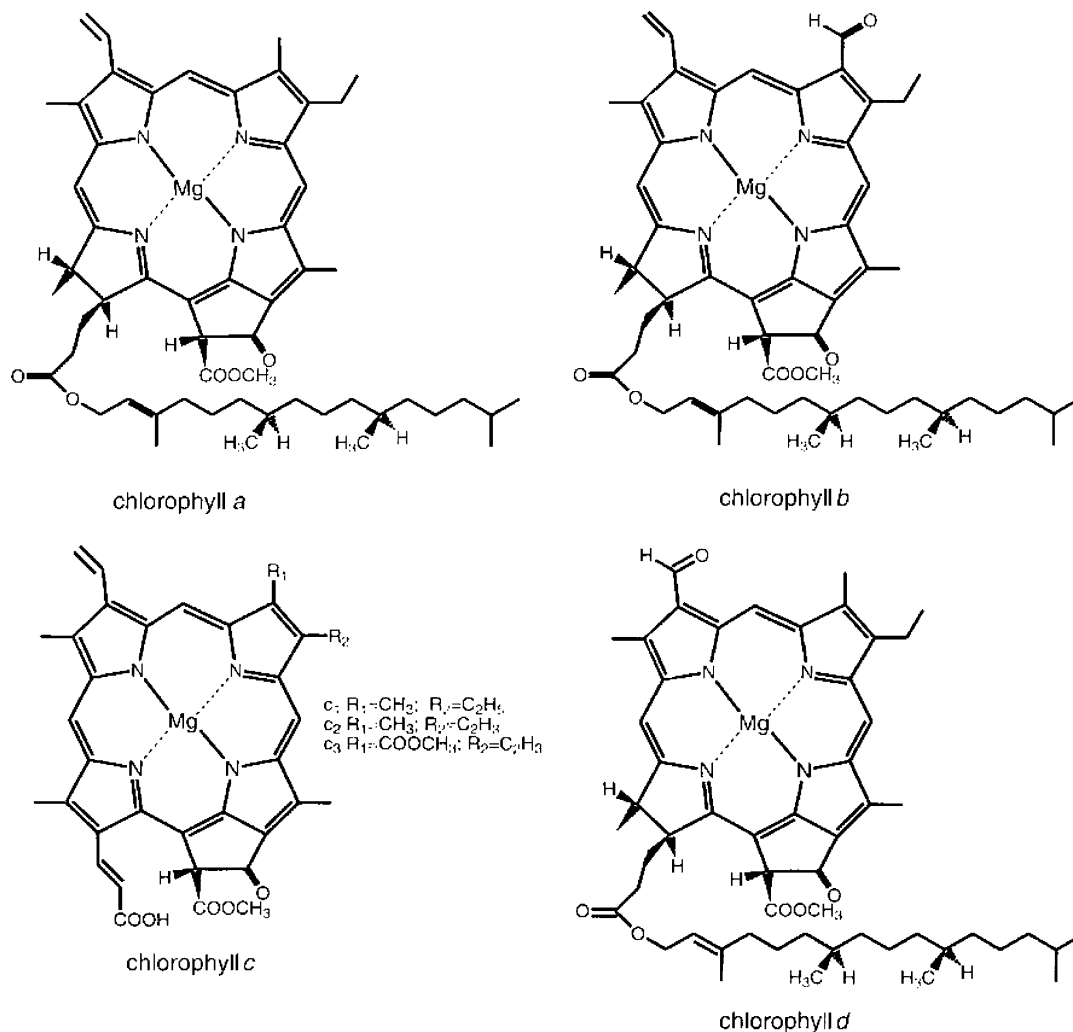


Figure 1. Structural formulas of the selected types of chlorophyll molecules.

the COH group is slightly less repelled by the surrounding protons and is better “incorporated” into the π -electron-conjugated system of the porphyrin ring. All lines in the spectrum of the cis conformer are mildly blue-shifted in comparison to those in the spectrum of the trans conformer (see Figure 6).

Electron Transition Spectra. For the global minima of the selected pigments, six electron singlet–singlet transitions were computed with the TDDFT method. In the discussion, usually three lines corresponding to visible (allowed) transitions are compared with experimental data. The full sets of transition lines are drawn in Figures 7–11 for chlorophyll molecules, bacteriochlorophylls, pheophytin, bilins, and carotenoids, respectively. Numerical values together with intensities and experimental data are collected in Tables 1–4. In the case of carotenoids, it is known from many studies (for instance, refs 54, 63, 74, and 75) that the TDDFT approach does not give the correct order for the S1 and S2 states. Therefore, TDA methods (as implemented in Q-Chem 3.0⁷⁶) were used with the same basis set and the non-hybrid functional BLYP.

Spectral characteristics

Chlorophylls. The spectra of chlorophyll molecules consist of two basic absorption bands. A most complex band, located in the blue region, is called the Soret band and is composed of several electron transitions. The other(s), called Q band(s), can usually be found in the red region. Their typical positions and intensities can be seen in the experimental spectra in Figure

12a. Such a distribution of electron transitions is responsible for the typical green or blue-green color of most chlorophyll pigments.

The calculated values of spectral lines are systematically blue-shifted in comparison to the experimental values by about 60–80 nm for the band Q_y and about 20–30 nm for the Q_x and Soret bands. Nevertheless, it is clear that a fairly good qualitative agreement with experimentally measured spectral lines has been obtained. In particular, the relative positions of spectral lines depending on various structural parameters (ligands) are in a very good accord with measured data for all chlorophyll types (both Chl and BChl).

Chlorophyll *a* is the most common chlorophyll pigment, and its spectrum is well established. The calculated values of the Q_y , Q_x , and B lines (B lines represent an edge of the Soret band) lie at 583, 539, and 398 nm, respectively. The measured intensity of the Q_y band (about 0.23) is in good accord with our results. However, we failed to reproduce the intensity of the Soret band of about 1.1 (similarly to some other researchers, e.g., ref 69). Nevertheless, in another measured spectrum (cf. Figure 12a), the intensities of both the Q_y and Soret bands are much closer to each other, in accord with our calculations.

From an analysis of molecular orbitals (MOs) and individual transitions it follows that the four-orbital model of Gouterman and co-workers can be still recognized as a reasonable approximation even within the TDDFT model. The first two transitions of chlorophylls (and also of bacteriochlorophylls and

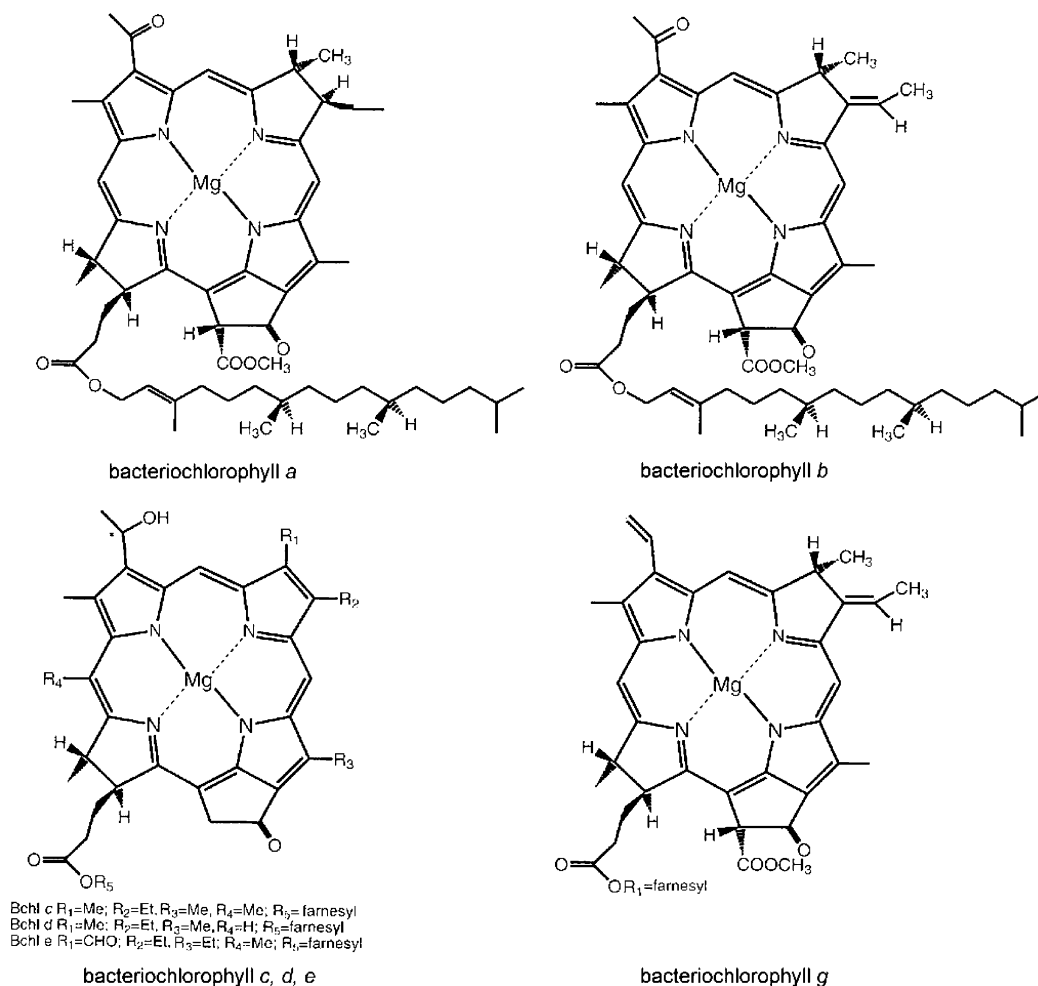


Figure 2. Structural formulas of the selected types of bacteriochlorophyll molecules.

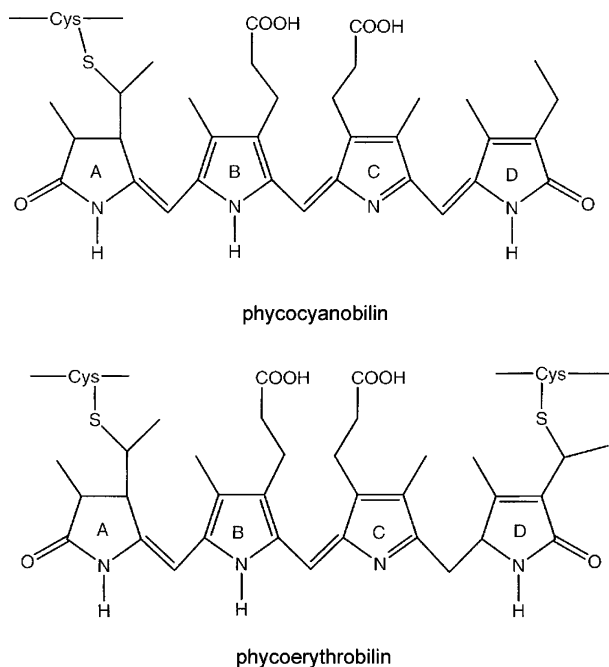


Figure 3. Structural formulas of the investigated phycobilins.

pheophytins) are always based primarily on these four MOs. The individual electron transitions can be compared with spectra of the Mg-porphin molecule.^{70,71} It can be noticed that HOMO has features of the a_{1u} HOMO - 1 orbital in symmetrical Mg-porphin and, inversely, HOMO - 1 has the character of a_{2u} of

the Mg-porphin HOMO (for the shapes of frontier orbitals, see Figure 13a-d). This can be generalized for all chlorophylls and bacteriochlorophylls with the exception of the Chl *c* family, in which the same order of frontier MOs as in Mg-porphin was found. However, these three systems have closer structural relationships to Mg-porphin than to the chlorin molecule (see below). Also, the near-degeneracy of the Mg-porphin HOMO and HOMO - 1 can be traced in the smallest energy difference of the HOMO and HOMO - 1 orbitals (less than 0.1 eV) in the case of chlorophyll *c*. The other frontier orbitals preserve the same character in all of the kinds of chlorophyll molecules considered herein. In structures without the D_{4h} symmetry of Mg-porphin, the LUMO and LUMO + 1 are no longer degenerate. This is in accord with the older theory, according to which such an MO crossing results from the removal of symmetry in passing from porphyrin to chlorin and bacteriochlorin models, as depicted in Figure 14. However, such a simple theory is not able to explain in detail the differences between individual types of chlorophylls and bacteriochlorophylls that are known from measured spectra and are computationally reproduced here.

Chlorophyll *b* differs from chlorophyll *a* in the C7 substituent group, where the methyl group of chlorophyll *a* is replaced by the formyl group.

In comparison with Chl *a*, a ligand variation causes a blue shift of the Q_y band of about 10 nm and a red shift of the Soret band of about 25 nm. From the orbital energies, it can be seen that the formyl group with the C=O double bond affects the conjugated system of chlorophyll *b*, decreasing the HOMO eigenvalue by about 0.3 eV. Otherwise, its character remains

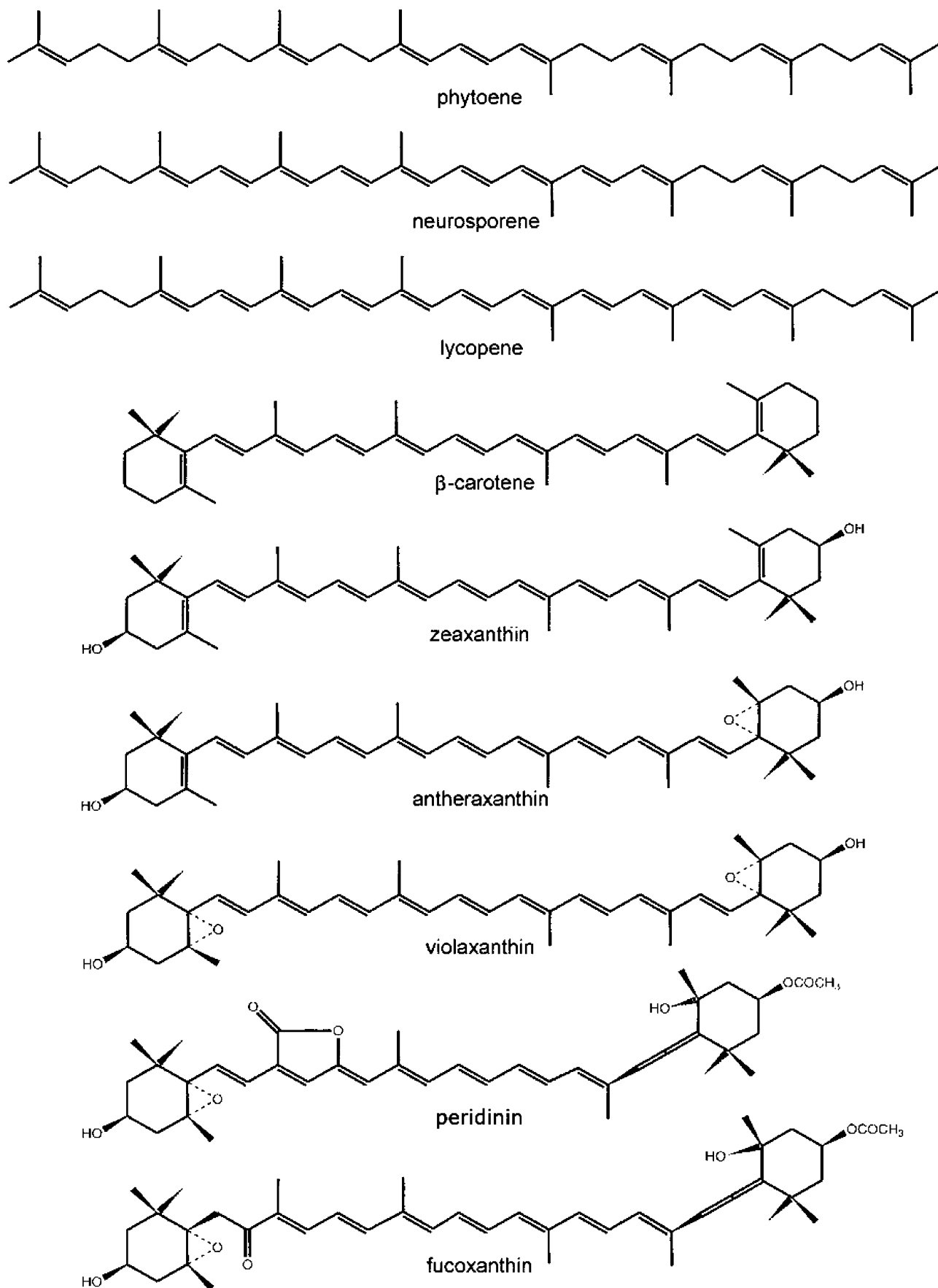


Figure 4. Structural formulas of the investigated carotenoids.

the same as in the Chl *a* molecule. However, the elongation of the π -conjugated system in the presence of the formyl group influences not only one MO. The $-\text{CHO}$ group causes an energy

shift of both the occupied and virtual frontier orbitals toward lower values, resulting in an actual blue shift of the first Q_y line with a reduced intensity. The red-shifted transitions in the

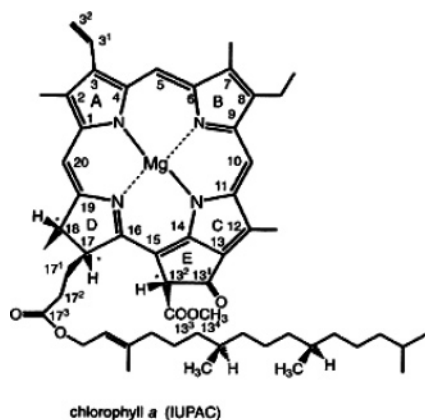


Figure 5. Chlorophyll *a*: atom numbering according to IUPAC conventions.

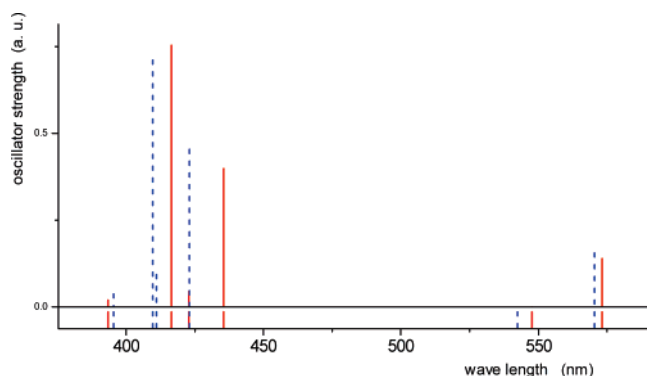


Figure 6. Lines of electron transitions of the chlorophyll *b* model molecules: positions of the COH group for the trans conformer (solid lines) and the cis conformer (dashed lines).

Soret band exhibit practically doubled intensities. The ratio of oscillator strengths of the Q_y band to the Soret band is markedly smaller for chlorophyll *b* than for chlorophyll *a*. This result is in very good agreement with the experimental data.⁷⁷ The analogous $-CHO$ group is also present in bacteriochlorophyll *e*, displaying a similar blue shift of the Q_y band within the BChl $c-e$ family.

Chlorophyll *c* molecules exhibit larger variability than the other chlorophylls. The isoprene tail, which is usually used for anchoring in the protein matrix, is not present. The pyrrole D ring is not reduced and thus the chlorophylls *c* more closely resemble the porphyrin system of the Mg-porphin molecule, whereas the other chlorophylls are derived from the chlorin system instead. This higher symmetry of the π -electron system explains why the spectral properties of chlorophylls *c* deviate from those of the other chlorophyll types. Also, as mentioned above, the order of the HOMO and HOMO -1 corresponds to the electronic structure of Mg-porphin. These modifications of chlorophylls *c* substantially influence the spectral transitions. They are responsible for stronger absorption in the Soret band (390–440 nm) and a reduction of the two Q bands in the visible region above 500 nm. The Q_y band is substantially reduced, and its intensity is almost eliminated.

Mutual differences between chlorophylls *c* correspond to different substituents on carbons C7 and C8, as can be seen from Figure 1. Whereas types *c1* and *c2* exhibit the most blue-shifted Q_y band of all of the chlorophyll systems, both ligands of chlorophyll *c3* have double bonds that interact with the π -electron system of the porphyrin ring, causing a visible red shift of the Q_y transition (by about 13 nm) in comparison to that of the *c1* and *c2* molecules. The energy of this spectral

line is very similar to that of the Q_y line of chlorophyll *a*. The Q_x transition lies relatively very close to the Q_y band for all three structures, less than 10 nm. This is again a consequence of the higher symmetry of the porphyrin ring.

Chlorophyll *d* differs from chlorophyll *a* in the C3 position, where the formyl group is present instead of the vinyl group. In this way, chlorophyll *d* can be considered as an intermediate between chlorophyll *a* and bacteriochlorophyll *a* (in terms of both spectral transitions and structure). The Q absorption lines exhibit a larger red shift than in any other chlorophyll. This is the only chlorophyll with a calculated Q_y transition over 600 nm. The explanation lies in the fact that the formyl group influences all frontier orbitals (both occupied and virtual), shifting their orbital energies to lower values. In particular, the eigenvalue of the LUMO displays the lowest energy among the whole set of examined chlorophylls. The π orbital of the formyl group is incorporated into the π -conjugated system of the porphyrin (or chlorin) ring where an elongation of the appropriate lobe is clearly visible (cf. Figure 13e). The same effect of the formyl group can be also seen in the case of bacteriochlorophylls *a* and *b*.

Bacteriochlorophylls. The electronic spectra of the bacteriochlorophylls are depicted in Figure 8 and in Table 2. The basic structural difference between chlorophyll and bacteriochlorophyll groups lies in the reduced B and D pyrrole rings (whereas in chlorophylls, only the D ring is reduced). Therefore, bacteriochlorophylls have a more perturbed π -electron system than chlorophylls. This should seemingly lead to the absorption of light with shorter wavelengths. However, this conclusion is not completely correct because the consequences of the perturbed conjugated ring are not so simple and straightforward. Moreover, the shortening of wavelengths can be observed only in the case of the Soret band. In contrast, the Q_y band displays a red shift in accordance with the experimental spectrum (cf. Figure 12b). The explanation of this red shift can be sought in the fact that, in the HOMO -1 (with a_{1u} character) of Mg-porphin, the p_z atomic orbitals (AOs) of carbons C7, C8, C17, and C18 (for atom numbering, see Figure 5) are involved. A reduction of these C7–C8 and C17–C18 bonds causes this MO to become increasingly less stable in passing to chlorin (or chlorophylls) and bacteriochlorin (or bacteriochlorophylls) (see Figure 14). Because the original HOMO (a_{2u}) of Mg-porphin is not affected by the reduction of the C–C bonds, the order of the HOMO and HOMO -1 is exchanged in chlorophyll and bacteriochlorophyll molecules. Similarly, only one of the original e_g virtual orbitals contains the p_z AO of the same carbon atoms (C7, C8, C17, C18). Thus, only this virtual orbital is destabilized in an analogous manner. Therefore, the LUMO is practically unaffected, whereas the energy of the LUMO $+1$ increases when a reduction of these two carbon bonds occurs (cf. Figure 14). This means that the gap between the HOMO and LUMO is actually smaller in bacteriochlorophyll than in chlorophyll systems, leading to Q_y absorption at longer wavelengths in bacteriochlorophylls.

It is interesting to notice that the position of the Q_x band is fairly stable, between 530 and 560 nm, regardless of whether chlorophylls, bacteriochlorophylls, or pheophytins are considered. Only the intensity is noticeably higher in the case of bacteriochlorophylls *a* and *b*.

All of the spectra of the bacteriochlorophylls calculated at the TD-B3PW91/6-31+G(d) level are collected in Figure 8. Generally, the computed transitions are blue-shifted from the measured absorption spectra, with an average shift of about 80 nm.

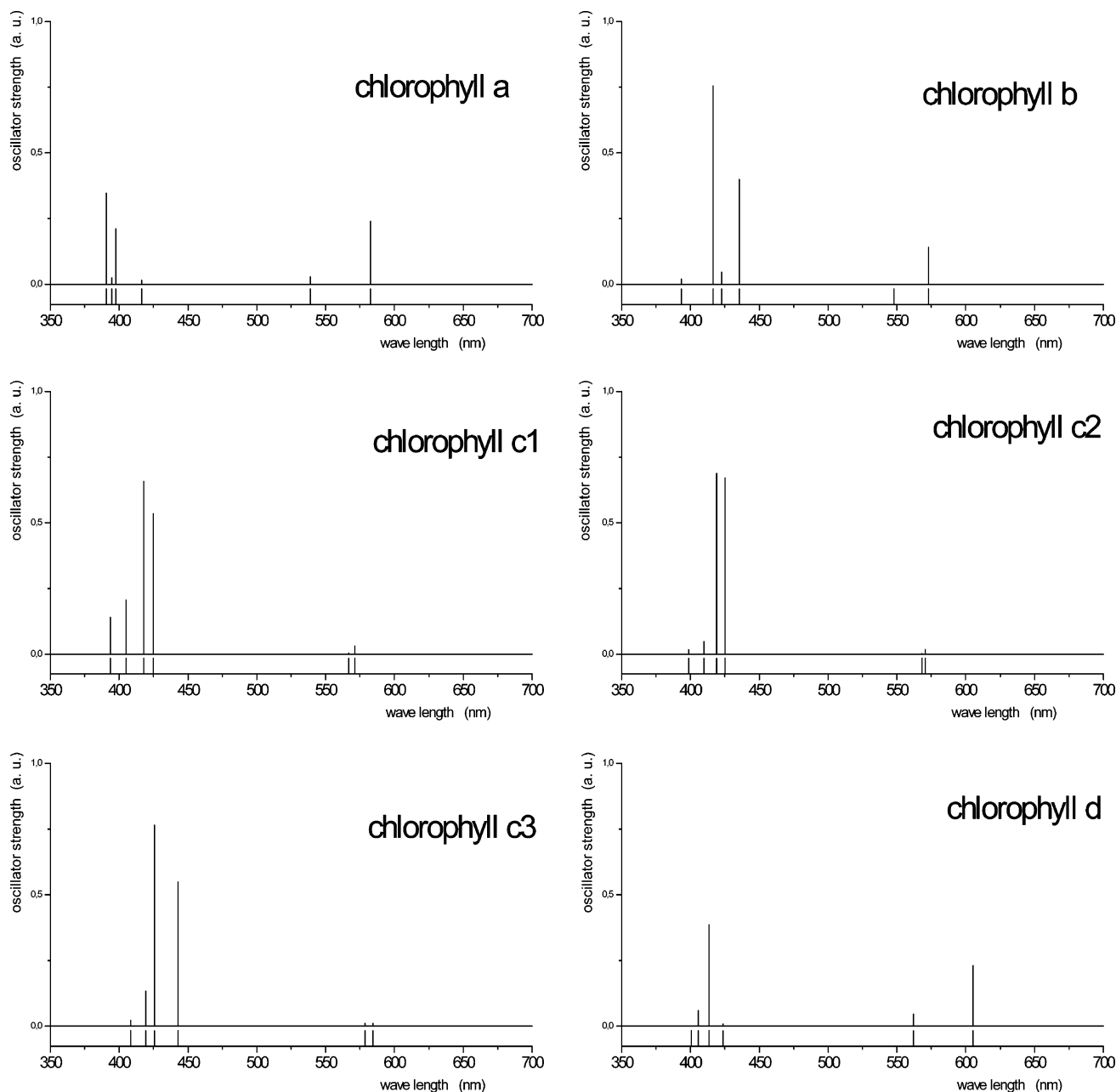


Figure 7. Computed (TDDFT) electron transitions of chlorophyll molecules.

Bacteriochlorophyll *a* is the most abundant chlorophyll-type pigment in the majority of anoxygenic photosynthetic bacteria and its spectrum is well-known (see Figure 12b). The increase of the intensity of the Q bands relative to the Soret band in comparison to the chlorophyll spectrum can be seen in both the measured (Figure 12) and calculated spectra (Figures 7 and 8).

A red shift of the Q_y band of more than 80 nm can be seen when transition energies are compared to the analogous spectra of chlorophyll *a*. Nevertheless, the position of the Soret band is practically the same (or slightly more blue-shifted). In accord with the four-orbital model, the calculated Q band is constituted predominately only by the HOMO \rightarrow LUMO transition; the contribution from the HOMO $- 1 \rightarrow$ LUMO $+ 1$ transition is markedly reduced (to the value of about 0.2) in comparison to that in chlorophylls (where this contribution is usually about 0.4). This is due to the increased gap of the unaffected orbital with a_{1u} character and the destabilized (because of the reduced C17–C18 bond) LUMO $+ 1$ with e_{gy} character. A similar

destabilization can also be observed in the HOMO with a_{2u} character (compare the energies of the HOMO at -0.188 au for BChl *a* vs -0.195 au for Chl *a*). Moreover, the LUMO with e_{gx} character is further stabilized by the presence of the formyl group, as also mentioned in the case of chlorophyll *d*. That is why the Q_y band is so strongly red-shifted.

Bacteriochlorophyll *b* contains an ethylidene ($=CH-CH_3$) group at the C8 site, which further extends the π -conjugated system (the HOMO energy is further decreased slightly to -0.185 au, preserving the eigenvalues of the other frontier orbitals), resulting in an additional red shift of the Q_y band. This pigment absorbs at the longest wavelength of any known chlorophyll type, i.e., according to our results, at about 700 nm (the experimental value is 794 nm). In comparison to BChl *a*, the Soret band is accordingly also red-shifted (the edge starts at about 400 nm). The B line is constituted mainly by the HOMO \rightarrow LUMO $+ 1$ electron transition.

Bacteriochlorophylls *c–e* form an “exceptional” group similar to chlorophylls *c*. The porphyrin ring is reduced only partially

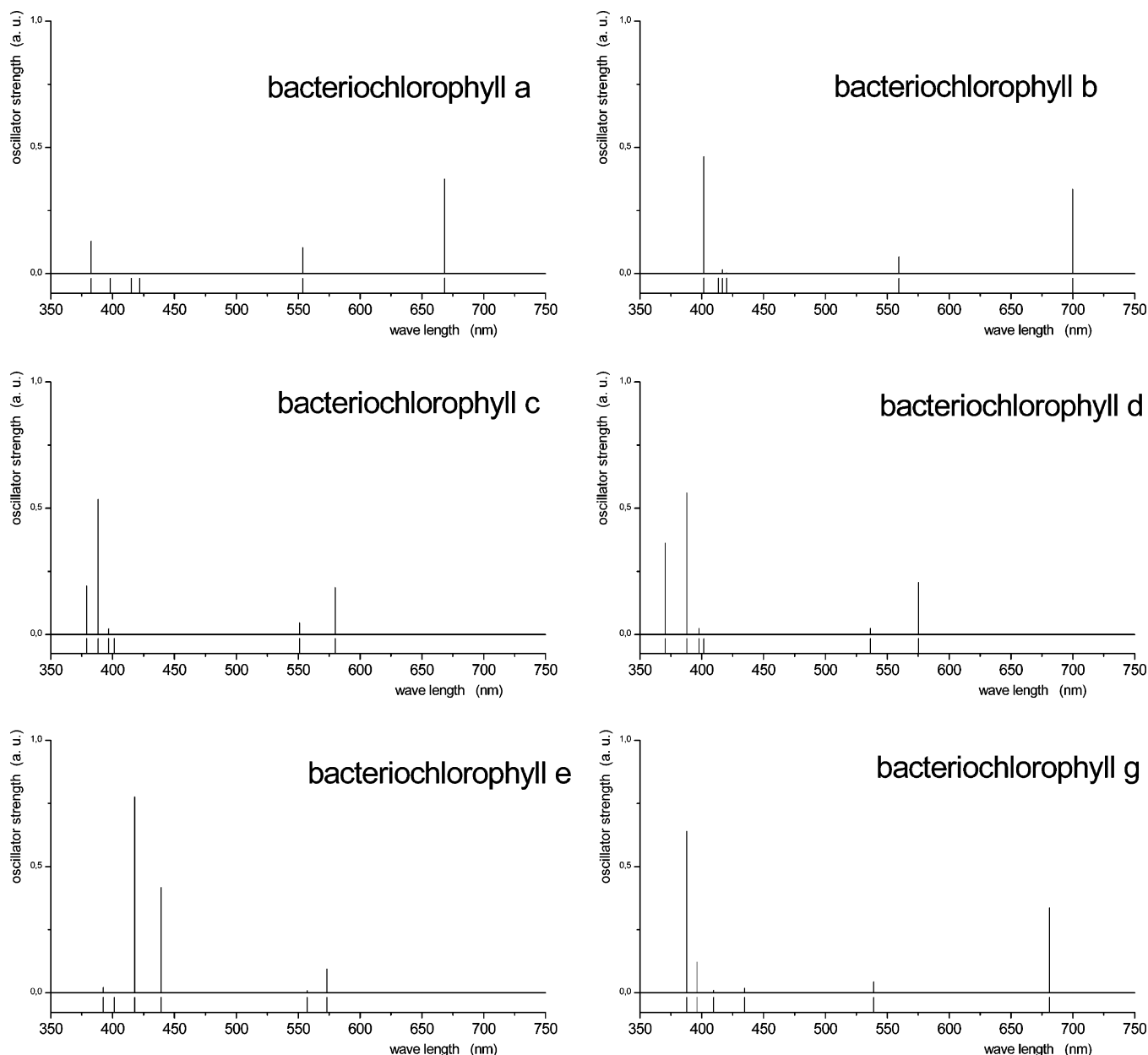


Figure 8. Computed (TDDFT) electron transitions of bacteriochlorophylls.

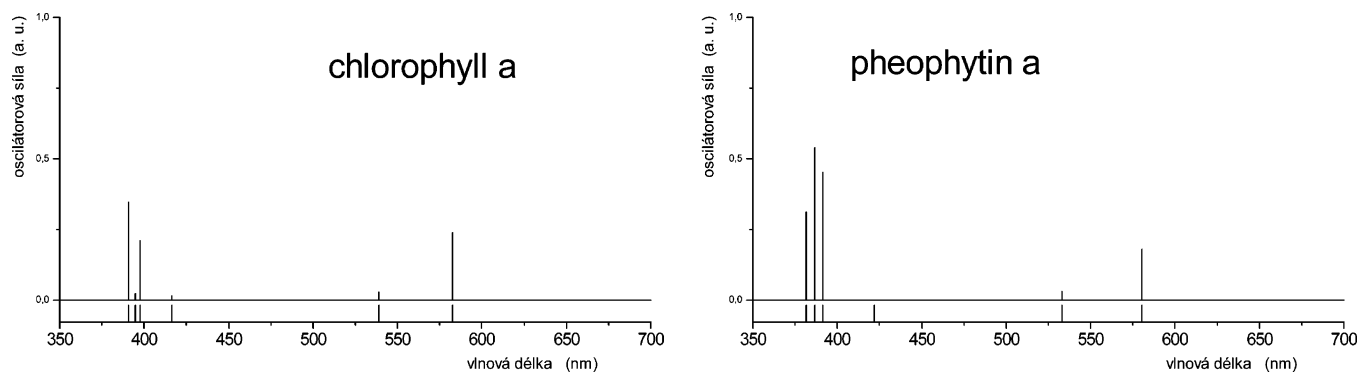


Figure 9. Comparison of calculated electron transitions of chlorophyll *a* and pheophytin *a*.

(C7–C8 bond), in the same way as in chlorophyll systems. Therefore, the structural and spectroscopic characteristics of these bacteriochlorophylls also resemble those of chlorophylls more closely. The Q_y band is positioned just below 580 nm and is composed of both the $HOMO \rightarrow LUMO$ and $HOMO - 1 \rightarrow LUMO + 1$ transitions with weights very similar to those

in the electron spectra of chlorophylls. Also, the intensities of the Q bands in relation to the Soret band are substantially lower.

The role of the formyl group in the case of BChl *e* can be clearly recognized from even more reduced Q_y intensity because the p_z AO of the formyl carbon is visibly present in the $HOMO$ and $LUMO + 1$ but not the $HOMO - 1$ and $LUMO$. Thus, the

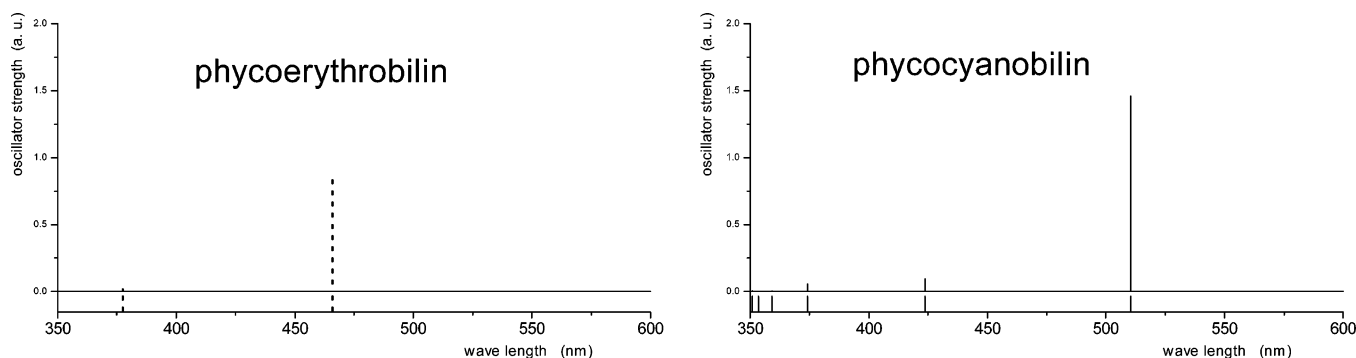


Figure 10. Computed (TDDFT) electron transitions of selected phycobilins. Phycoerythrobilin: dotted lines for the complex with one cysteine molecule and solid lines for the pigment with two cysteine molecules.

overlap of the relevant orbitals in the transition dipole moment formula is decreased. The additional small blue shift of the first transition line is related to a larger stabilization of the HOMO again as a consequence of the incorporation of the formyl p_z AO into the π -conjugated system.

Some changes also occur in the Soret band, which is formed predominately by the HOMO \rightarrow LUMO + 1 transition. Because the formyl p_z AO is involved in both of these MOs, the intensity of this band is increased accordingly. An apparent shift toward longer wavelength can be noticed. Moreover, at the edge of the Soret band, another transition appears in which the HOMO - 1 \rightarrow LUMO + 1 transition dominates (its character is based on an inverse combination of the excitation weights of the Q_y line). All of the energies and intensities are in agreement with the experimental data obtained for all three bacteriochlorophylls.⁷⁷

Bacteriochlorophyll *g* has a vinyl group located at the C3 position as in chlorophyll *a*. Moreover, an ethylidene group is present at the C8 site. Bacteriochlorophyll *g* was found to be relatively unstable; it can be easily reduced by saturation of the ethylidene double bond and converted into chlorophyll *a*. In its spectrum, the Q_y band is positioned at about 680 nm with a prevailing HOMO \rightarrow LUMO transition similar to Q_y of bacteriochlorophyll *a*. Comparing Q_y with the analogous band of chlorophyll *a*, a red shift of about 98 nm can be noticed, in accord with an analogous experimental difference of ca. 105 nm. The edge of the Soret band is also made up of two transitions. In addition to the usual HOMO - 1 \rightarrow LUMO line (at 388 nm), another less intense line at 396 nm can be observed.

Pheophytin *a*. All chlorophylls and bacteriochlorophylls have their own pheophytin or bacteriopheophytin variants. Structurally, the only difference lies in the absence of the magnesium cation in the center of the porphyrin ring. In this study, the electron spectrum of only the pheophytin *a* molecule was explored for the purpose of comparison with the most common chlorophyll *a*. In Figure 9 and Table 1, the spectrum of pheophytin *a* is presented. The spectra of chlorophyll *a* and pheophytin *a* exhibit prominent similarities. Small but significant differences are in good agreement with analogous differences found in the measured spectra in Figure 12a. The Q transitions are more intense in chlorophyll than in pheophytin. The oscillator strengths for the first and second transitions of chlorophyll are more than 30% larger. Similar results were also published by Hasegawa et al.,²⁵ Parusel and Grime,²⁶ and Sundholm.³³ Because the AOs of the magnesium cation are not involved in any of the frontier MOs of the π -conjugated system, the small differences between the chlorophyll and pheophytin spectra can be readily understood. Also, an approximately 2 times larger intensity of the Soret band with respect to the Q_y band was obtained for pheophytin, which is in agreement with

the experimental spectra in Figure 12a. Our calculations failed to describe the relative difference of the Q_x position compared to the Q_y band, which is about 30 nm in the measured spectra but less than 10 nm in the TDDFT results (see Table 1).

Phycobilins. These pigments can be characterized by open-chain tetrapyrrole structures covalently bound to proteins (cf. Figure 3). Their structures are very flexible, giving fairly different spectral bands. This feature is partially described in several textbooks, such as ref 78. Therefore, the relevant “biological structures” were taken from the Protein Data Bank.

Phycoerythrobilin and phycocyanobilin were chosen for a study of the influence of structure on spectral properties. The basic difference between these two molecules lies in the chromophore length. Whereas phycoerythrobilin has one saturated bridge ($-\text{CH}_2-$) and two unsaturated bridges ($-\text{CH}=\text{}$), phycocyanobilin has all bridges unsaturated. A longer conjugated chromophore chain in phycocyanobilin is clearly visible in the spectral differences between the two molecules (cf. Figure 10 and Table 3), in accordance with the experimental results depicted in Figure 15. However, the unsaturated chromophore chain is far from being planar, and thus, the π -conjugated system is severely perturbed, as can be seen in Figure 16 where the optimized structure of phycoerythrobilin is shown. From Figure 3, another difference can be noticed, namely, the number of bonds to proteins. Whereas phycocyanobilin is attached to the peptide through a single covalent bond, in the case of phycoerythrobilin, both ends are fixed in the protein matrix. However, this fact has no substantial effect on the electronic spectra. We also examined the single-anchoring phycoerythrobilin. It was confirmed that the number of anchoring bonds has no substantial influence on the transition spectrum, as can be seen in Figure 10 and Table 3. The predicted spectral lines are in good agreement with the measured data. The estimated position of the first line is blue-shifted by about 100 nm from the experimental values. Comparing the two phycobilins, the relative positions of spectral lines also agree well with experiment. The first spectral line in phycoerythrobilin is blue-shifted by about 44 nm from the corresponding line of phycocyanobilin, in fairly good accord with the experimental shift of 53 nm. The intensities of these lines are not of great concern because they depend strongly on the structure. Moreover, the effect of the environment is also more important in these molecular systems.

Carotenoids. The common feature of carotenoids is a linear chain with a varying length of unsaturated conjugation. Carotenoids have relatively simple absorption spectra. They usually have three absorption bands corresponding to transfers from the lowest vibration level of the ground state to the three lowest vibration levels of the first excited state.⁷⁹ Because the vibration problem was not considered in our study, only the position of the first spectral line can be compared to experimental data.

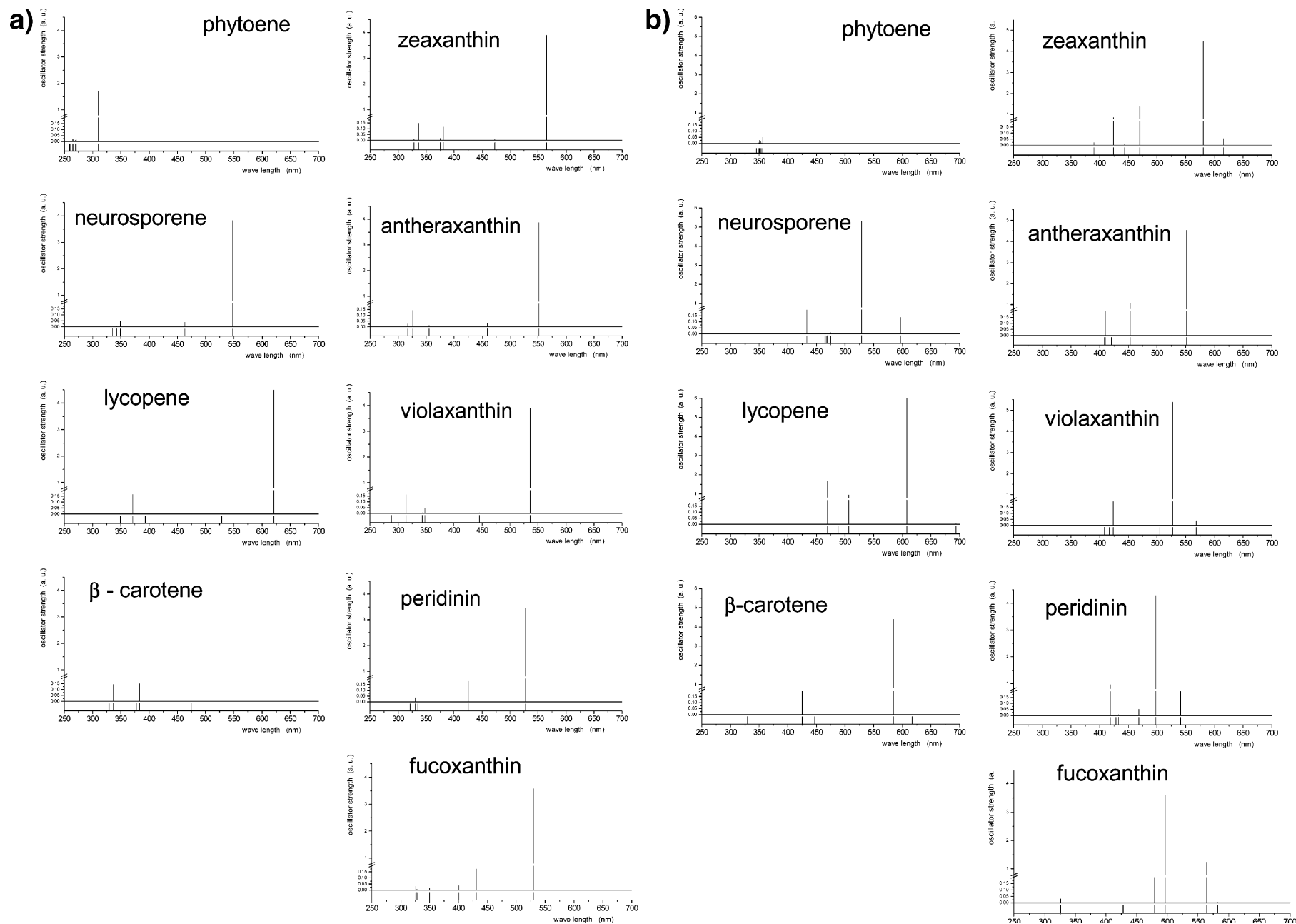


Figure 11. (a) TDDFT and (b) TDA electron transitions of the most frequent carotenoids.

TABLE 1: Experimental Data and Computed (TDDFT) Electron Transitions for Chosen Chlorophyll Structures^a

Chl a	λ [nm] in diethyl ether	662	578	430		
	λ [nm]	583	539	416	398	395 391
	oscillator strength	0.24	0.03	0.02	0.21	0.03 0.35
Chl b	λ [nm] in diethyl ether	644	549	455		
	λ [nm]	573	548	436	423	417 393
	oscillator strength	0.14	0.	0.40	0.05	0.75 0.02
Chl c1	λ [nm] in diethyl ether	628	578	444		
	λ [nm]	571	567	425	418	405 394
	oscillator strength	0.03	0.	0.54	0.66	0.21 0.14
Chl c2	λ [nm] in diethyl ether	628	579	448		
	λ [nm]	571	568	425	419	410 399
	oscillator strength	0.02	0.	0.67	0.69	0.05 0.02
Chl c3	λ [nm] in diethyl ether	626	586	452		
	λ [nm]	584	579	443	426	419 408
	oscillator strength	0.01	0.01	0.55	0.77	0.13 0.02
Chl d	λ [nm] in diethyl ether	668	-	447		
	λ [nm]	605	562	424	413	406 401
	oscillator strength	0.23	0.05	0.01	0.38	0.06 0.
Pheo a	λ [nm] in diethyl ether	672	540	416		
	λ [nm]	580	533	422	391	387 381
	oscillator strength	0.18	0.03	0.	0.45	0.54 0.31

^a Experimental data on gray background taken from Scheer.²**TABLE 2: Experimental Data and Computed (TDDFT) Electron Transitions for Bacteriochlorophyll Molecules^a**

BChl a	λ [nm] in diethyl ether	773	577			358
	λ [nm]	668	554	422	415	398 383
	oscillator strength	0.37	0.10	0.	0.	0. 0.13
BChl b	λ [nm] in diethyl ether	794	578			368
	λ [nm]	700	559	420	417	414 402
	oscillator strength	0.33	0.07	0.	0.02	0. 0.46
BChl c	λ [nm] in diethyl ether	659	-			429
	λ [nm]	580	551	401	397	388 379
	oscillator strength	0.19	0.05	0.	0.02	0.54 0.19
BChl d	λ [nm] in diethyl ether	651				423
	λ [nm]	575	536	402	398	388 370
	oscillator strength	0.21	0.02	0.	0.03	0.56 0.36
BChl e	λ [nm] in diethyl ether	647	-	458		
	λ [nm]	573	557	439	418	401 392
	oscillator strength	0.09	0.01	0.42	0.78	0. 0.02
BChl g	λ [nm] in diethyl ether	767	565			404
	λ [nm]	681	539	435	409	396 388
	oscillator strength	0.34	0.04	0.02	0.01	0.12 0.64

^a Experimental results on gray background taken from Scheer.²**TABLE 3: Experimental Data^a and TDDFT Computed Electron Transitions for Phycobilins**

phycocyanobilin	water	617	555			
	λ [nm]	510	424	374	359	354 351
	oscillator strength	1.46	0.10	0.06	0.	0. 0.
phycoerythrobilin (1AA)	water	565	495			
	λ [nm]	466	390	377	348	333 326
	oscillator strength	0.86	0.	0.02	0.01	0.01 0.01
phycoerythrobilin (2AA)	λ [nm]	466	378	349	333	332 326
	oscillator strength	0.83	0.02	0.	0.02	0. 0.05

^a Results on gray background were taken from the Prozyme Phycobiliproteins Web page.⁸¹

In the computational model, only all-trans isomers were considered. All obtained spectra exhibit a systematic red shift relative to the experimental data that is roughly proportional to the number of unsaturated double bonds. For the reddest line, the shift varies from 13 nm, for phytoene with three double bonds, to 119 nm, for lycopene with a π -conjugated system formed by 11 double bonds. The average shift is about 70 nm. This variation is an unpleasant property and makes the spectral

TABLE 4: Experimental Data^a and Computed Electron Transitions at the TDA/TDDFT(BLYP)/6-31+G* and TDDFT(B3PW91)/6-31+G* Levels for Carotenoids

phytoene	λ [nm] in hexan	297	286	276		
	TDA λ [nm]	356	354	352	350	349
	oscillator strength	0.05	0.	0.02	0.02	0.
full	λ [nm]	310	271	269	265	260
	oscillator strength	1.71	0.01	0.01	0.02	0.
neurosporene	λ [nm] in hexan	468	440	415		
	TDA λ [nm]	597	529	475	468	465
	oscillator strength	0.14	5.31	0.01	0.	0.01
full	λ [nm]	548	463	355	349	343
	oscillator strength	3.83	0.04	0.08	0.05	0.
lycopene	λ [nm] in hexan	502	470	444		
	TDA λ [nm]	694	608	506	488	488
	oscillator strength	0.01	6.07	0.94	0.	0.
full	λ [nm]	621	528	409	393	371
	oscillator strength	4.50	0.	0.11	0.	0.16
β-carotene	λ [nm] in hexan	478	450	425		
	TDA λ [nm]	617	585	470	447	425
	oscillator strength	0.	4.38	1.54	0.	0.79
full	λ [nm]	566	474	383	377	337
	oscillator strength	3.87	0.	0.15	0.	0.14
zeaxanthin	λ [nm] in ethanol	478	450	425		
	TDA λ [nm]	615	581	470	444	424
	oscillator strength	0.05	4.46	1.37	0.01	0.85
full	λ [nm]	565	473	381	375	336
	oscillator strength	3.89	0.	0.11	0.02	0.15
antheraxanthin	λ [nm] in ethanol	474	445	422		
	TDA λ [nm]	596	552	453	421	410
	oscillator strength	0.33	4.52	1.06	0.	0.25
full	λ [nm]	551	459	371	355	326
	oscillator strength	3.85	0.08	0.09	0.01	0.14
violaxanthin	λ [nm] in ethanol	470	440	419		
	TDA λ [nm]	568	526	505	424	417
	oscillator strength	0.08	5.37	0.	0.53	0.
full	λ [nm]	536	445	348	343	314
	oscillator strength	3.89	0.01	0.05	0.	0.16
fucoxanthin	λ [nm] in ethanol	475	449	426		
	TDA λ [nm]	582	565	496	479	428
	oscillator strength	0.	1.24	3.59	0.71	0.
full	λ [nm]	529	431	400	350	328
	oscillator strength	3.57	0.17	0.04	0.02	0.
peridinin	λ [nm] in hexan	485	455			
	TDA λ [nm]	541	498	468	433	429
	oscillator strength	0.55	4.28	0.05	0.	0.
full	λ [nm]	528	426	350	335	331
	oscillator strength	3.45	0.18	0.05	0.	0.04

^a Results on gray background were taken from Carotenoids.⁷⁹

analysis somewhat more complicated than in the previous case of tetrapyrrole systems. Nevertheless, comparing the relative positions of the spectral lines, acceptable agreement with the measured spectra is achieved. The electron transition lines for the complete set of chosen carotenoids are presented in Figure 11 and Table 4. In the TDA spectrum of phytoene, the molecule with the shortest π -conjugated system, some problems can be noticed. Whereas a pronounced intensity for the excitation to S2 state can be seen in the full TDDFT(B3PW91) results, in the results obtained by the TDA method, the intensity of this transition is too low. However, a very similar spectrum was also obtained at the TDDFT(BLYP) level, demonstrating that this particular problem is instead related to the BLYP functional.

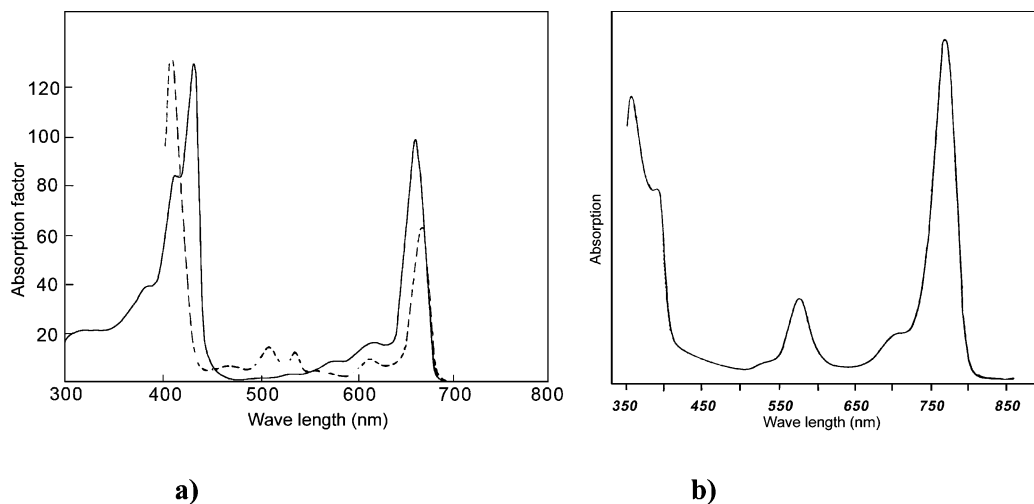


Figure 12. Experimental absorption spectra of (a) chlorophyll *a* (—) and pheophytin *a* (---) and (b) bacteriochlorophyll *a*. All spectra measured in diethyl ether.²

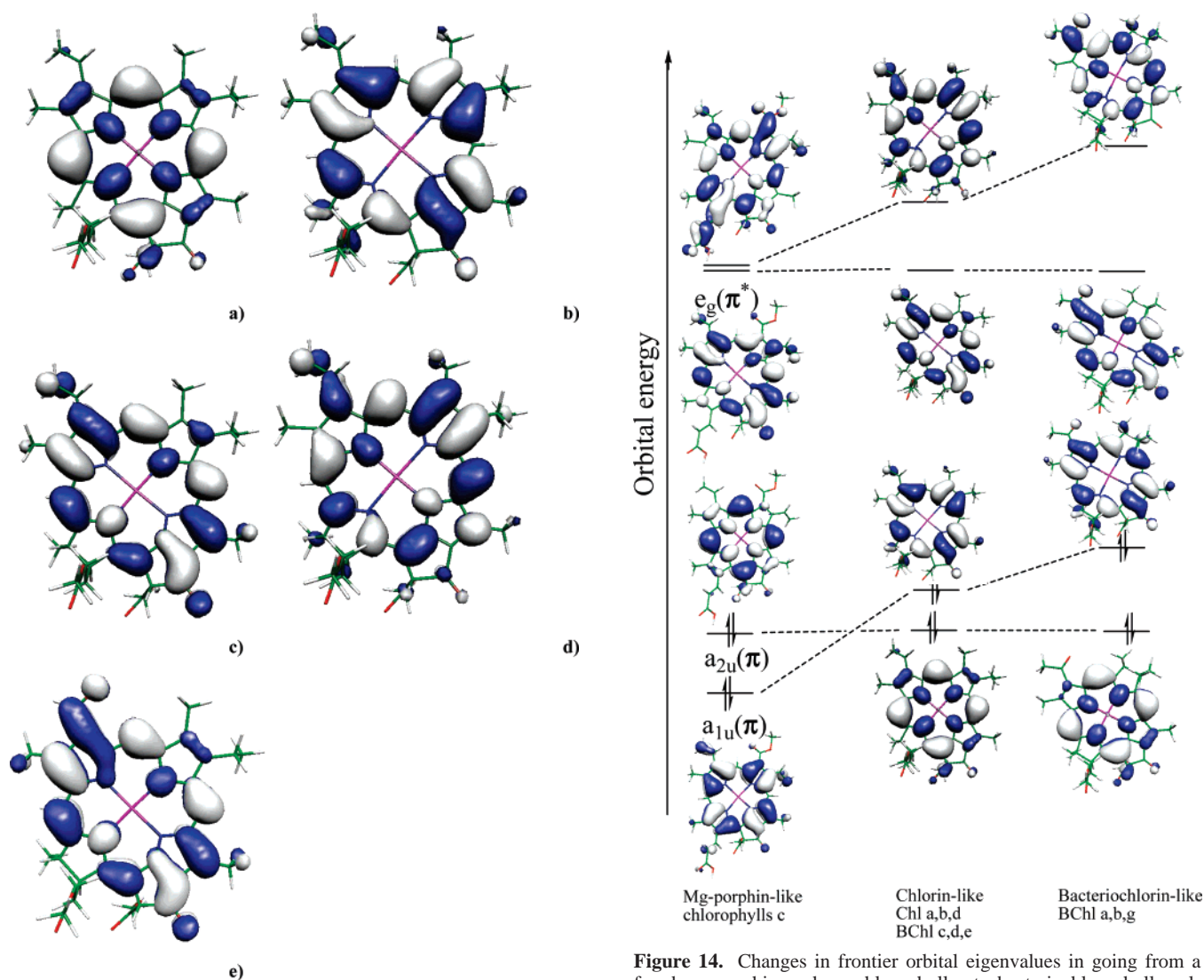


Figure 13. (a–d) Frontier MOs of chlorophyll *a*: MOs from 165 to 168; the 166th MO is the HOMO. (e) LUMO of chlorophyll *d*.

S1 and S2 States. In all determined spectra, a remarkably strong absorption (with high intensity) of the first visible transition was found. This line is related to the excitation from the HOMO to the LUMO, which have opposite symmetry according to the classification of unsaturated chains within

Figure 14. Changes in frontier orbital eigenvalues in going from a free-base porphyrin such as chlorophyll *c* to bacteriochlorophyll molecules.

Hückel theory. According to this concept, “gerade” and “ungerade” MOs alternate regularly. This feature is largely preserved in more accurate calculations as well. Therefore, large transition dipole moments can be expected for the excitation between the HOMO and LUMO, resulting in a high intensity

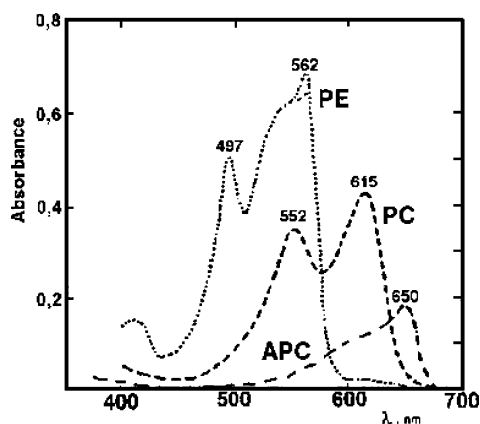


Figure 15. Absorption spectra of phycoerythrobilin (PE), phycocyanobilin (PC), and allophycocyanin (APC).

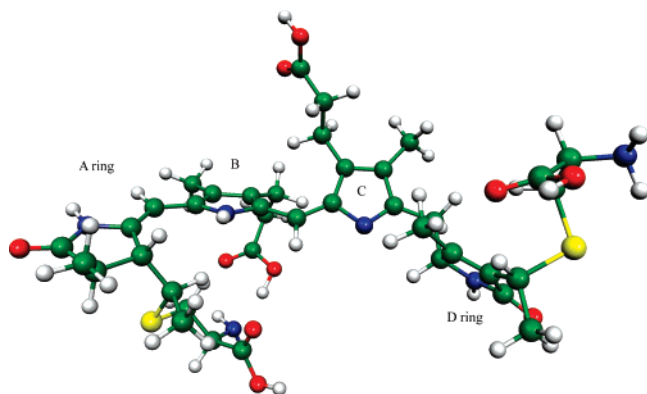


Figure 16. B3PW91/6-31G(d)-optimized structure of phycoerythrobilin anchored to two cysteine molecules.

of this transition. The second singlet $^1B_u^+$ excitation is predicted as the first one at the TDDFT level, whereas the TDA approach correctly inverts the order of the excited states, placing the $S1(2^1A_g^-)$ excitation at longer wavelengths. The A_g^- line is associated with an electron transition from the HOMO - 1 to the LUMO with some admixture of the HOMO \rightarrow LUMO + 1 transition. This excitation has a very low intensity (invisible in spectra), which again follows from simple Hückel theory. Because the two MOs have the same gerade symmetry, there is always a very small transition dipole moment between such orbitals. When the character of the unsaturated π -conjugated system is substantially perturbed, as in peridinin, fucoxanthin, and antheraxanthin, an increase in the intensity of this spectral line can be noticed.

Effect of the Chromophore Chain Length. As the chromophore chain elongates, electrons are increasingly delocalized, decreasing the excitation energy. Therefore, the increasing length of the conjugated chain strongly correlates with a longer wavelength for electron transitions in the following order: phytoene, neurosporene, and lycopene, i.e., in carotenoids without any other structural features such as β -cycles. In the case of phytoene, only three conjugated double bonds are present (with the first absorption at about 310 nm at the TDDFT level); neurosporene has nine (about 550 nm) and lycopene has 11 conjugated double bonds (with excitation energy of about 620 nm). Also, the intensity of the line increases with the elongation of the chromophore chain. The extension of the π -conjugated chain by two double bonds causes an increase in the wavelength of about 80 nm. This is slightly larger than the corresponding experimental shift (62 nm). As mentioned above, the length of

the conjugated system influences the size of the deviation between the computational and experimental values.

Effect of β -Cycles. Lycopene and β -carotene contain π -conjugated systems of the same length. However, in the carotene molecule, two additional β -cycles are present on the two ends of the linear chain where the edge double bond is incorporated. Comparing the two carotenoids, the effect of the β -cycle can be elucidated. It causes a blue shift of the S2 transition lines from 608 (621) nm according to the TDA (TDDFT) method for lycopene to 585 (566) nm for β -carotene. Similar shifts can be seen for other spectral lines of β -carotene as well. A smaller oscillator strength of this transition is visible in for this molecule. The experimentally observed difference between lycopene and β -carotene in the first absorption peak is practically the same (24 nm) as predicted by the TDA methods. On the other hand, the question of whether β -carotene should still be considered as a system with 11 conjugated double bonds remains in dispute because two of the bonds (one at each end) are involved in β -cycles. Because these cycles deviate from the plane of the π system, they represent a perturbation of the conjugation. Moreover, the spectrum of β -carotene is closer to the spectrum of neurosporene, in terms of not only the frequencies of the lines but also the intensity of the first visible transition.

Effect of the C3 Hydroxyl Group. Another comparison involves zeaxanthin and β -carotene. The zeaxanthin molecule is a derivate of β -carotene in which two hydroxyl groups are located on the C3 carbons of the β -cycles. The experimental spectra of these two pigments are practically identical. This means that hydroxyl groups have no substantial effect on the spectrum. This can be explained by the fact that a hydroxyl group cannot be incorporated into π -electron conjugation. This was also confirmed by the analysis of frontier MOs. Practically no expansion coefficients for the AOs of oxygen were detected within the five highest-occupied and four lowest-energy virtual MOs. The estimated electron transitions are in fairly good accord with the experimental spectra.

Effect of Epoxy Groups and Xanthophyll Cycles. The formation of xanthophyll cycles is an important feature present in regulation processes that control the wavelength of light absorption (rather than the intensity). In Figure 11, the blue shift of the S2 transition is visible passing from zeaxanthin through antheraxanthin to violaxanthin. The shift is nearly equidistant in both the experimental and calculated values and originates from the replacement of the double bond present in the β -cycle by an oxygen insertion to form an epoxy group. This causes a shortening of the π -conjugated system by one double bond on each side of the chain. Nevertheless, the change in energy of the first allowed transition (S2), about 25 (15) nm by TDA (TDDFT), is not fully comparable to the above-discussed effect of the length of the chromophore chain because these double bonds are already involved in β -cycles. Nevertheless, the computed energy differences for transition to the S2 state between the three xanthines are still substantially larger than the corresponding experimental values as a consequence of the overestimated dependence of the transition lines on the length of the π -conjugated system (as mentioned above). The intensity of the S2 transition increases during the epoxidation from zeaxanthin to violaxanthin at the TDA level, in contrast to the TDDFT results, according to which the intensity remains practically constant. The red shift of the S2 spectral line, going from violaxanthin to zeaxanthin, could be related to the possibility of starting a nonphotochemical quenching (NPQ) mechanism, expecting that the de-epoxidized zeaxanthin forms an "energy sink" as suggested in ref 1. However, another

possibility of NPQ should be stressed. The TDA results exhibit a larger difference between the S1 and S2 states in the violaxanthin spectrum (about 60 nm) than in the zeaxanthin spectrum, where only a 40 nm gap was found. This means that, in zeaxanthin, the S1 and S2 energy levels are substantially closer, which can clearly play a key role in NPQ mechanism.

Effect of Other Groups. Peridinin and fucoxanthin can be used to elucidate another structural feature of carotenoids. Both contain the allenic group ($-\text{CH}=\text{C}=\text{CH}-$). The presence of this group plays an important role in plants such as algae that do not contain chlorophyll *b* in their antenna complexes. In addition, fucoxanthin contains one carbonyl group at the C8 position, and peridinin contains a lactone group near the C8 position, which makes the analysis of various structural groups more complicated. The carbonyl group together with the allenic moiety makes the spectrum slightly red-shifted compared to the spectrum of violaxanthin. From Table 4, it can be seen that the experimental red shift of 5 nm qualitatively differs from the computed blue shift of 7 nm (using the TDDFT method) or 30 nm (at the TDA level). However, it can be guessed that the relative error in the estimation of spectral lines is more than 10 nm. In the TDA-predicted spectra, another state (1^1B_u^-) was localized between 1^1B_u^+ and 2^1A_g^- . This state is mainly made up of the HOMO $- 2 \rightarrow$ LUMO transition and lies slightly higher in other carotenoids.

The experimental spectral bands of peridinin are slightly red-shifted in comparison to the corresponding bands of fucoxanthin (by about 10 nm). The calculated results obtained by TDA match this trend well. From the comparison of the similar spectra of fucoxanthin and peridinin, the superior performance of TDA over the full TDDFT method is apparent in this particular case of carotenoids. Some additional support for the better performance of TDA can be also seen in the work of Dreuw et al. on increasing the length of polyene chains, in which results from more sophisticated ADC(2) (algebraic diagrammatic constructions to the second order) are compared.⁷⁴ The TDDFT results do not reproduce the relative positions of peridinin spectral lines, whereas the TDA approach fits the experimental values substantially better.

Error Estimation of the Electron Transitions. The predicted first spectral lines of (bacterio)chlorophyll and phycobilin molecules are usually shifted by up to 100 nm toward longer wavelengths in comparison to the experimental results. Smaller differences can be seen for higher excitation energies. In the blue region, the error usually drops to below 40 nm. The differences between measured and computed lines vary for every group of pigments. However, the effect caused by various ligands within a group of pigments is easily visible. It can be mentioned that the error in the spectral line determination is inherent in the chosen method, basis set, and system (here, π -conjugated systems of tetrapyrrole structures with similar sizes) and is therefore practically constant for these particular pigment groups. Another source of deviations between experimental spectra and calculated lines is related to different environments or surrounding molecules. Our calculations at this stage were performed for the in vacuo model. Considering these aspects, a very good accuracy (small relative error for the each given spectral line) with experimental data was achieved for the TDDFT(B3LYP) method and the chosen basis set.

In contrast to chlorophyll/phycobilin systems, the determination of the electron absorption spectra of carotenoids is related to a fundamental problem. That is, the error in spectra determination is dependent on the examined systems and exhibits a dependence on the size of the π -conjugated system,

as mentioned above. Therefore, it is more difficult to estimate even relative positions of various spectral lines.

Conclusion

In this study, various pigments from photosystems were explored. The most stable conformations were chosen based on a database search, with subsequent geometry optimization performed in several steps at different computational levels.

For the global minima, electronic excitations were determined using the TDDFT method at the B3PW91/6-31+G(d) level. For the group of carotenoids, the TDA approach with the non-hybrid BLYP functional and the same basis set was compared.

In the spectra of chlorophylls and pheophytin transition lines, lower intensities were determined for the Q bands than for the Soret band. Spectral energies of the Q_y lines were systematically blue-shifted by about 50–80 nm. Nevertheless, the correct order of these Q lines among various chlorophyll types can be noticed from a comparison with experimental results. A much better agreement was obtained for the edge of the Soret band represented by B transitions, for which the difference was at most 35 nm in the case of chlorophyll *d*.

Similar conclusions also hold for bacteriochlorophylls. The Q_y line is slightly more shifted, on average, in comparison to the experimental transition—by about 85 nm. For the bacteriochlorophylls, a correct higher intensity of the Q_y band compared to the Soret band was obtained.

In the case of phycobilins, the large flexibility of the open tetrapyrrole system can lead to quite artificial electron transitions. Therefore, obtaining the proper molecular conformation from the structural database was very important. The first transition line of phycobilins occurs at substantially lower wavelengths (by around 500 nm) and has a dominant character (high intensity). The influence of anchoring a cysteine side chain was found to be relatively small.

Spectral lines of carotenoids are based on linear polyene chains, which have a dominant influence on the spectra. Nevertheless, other structural elements such as β -cycles and epoxy, carbonyl, and allenic groups were also investigated. It was found that the first allowed excitation to the S2(1^1B_u^+) state has a very high intensity because of the different characters of the HOMO and LUMO. According to simple Hückel theory, the even and odd characters of the MOs alternate. Therefore, a large transition dipole moment can be expected for this spectral line. In contrast, the transition to S1(2^1A_g^-) is, according to the same arguments, forbidden. Because simple Hückel theory is not completely valid (carotenoids do not exhibit the symmetry of the simplified polyene model), some small intensities can be noticed. The largest values appear in the cases of peridinin, fucoxanthin, and antheraxanthin, where the influence of allenic and xanthophyll groups is notable.

In the case of fucoxanthin, the 1^1B_u^- state was found to lie between the 2^1A_g and 1^1B_u^+ states; it usually lies above this state in the other investigated carotenoids.

In contrast to the first visible transition lines of chlorophyll molecules, in the case of carotenoids, the S2 transition lines are overestimated (red-shifted) by about 70 nm on average. Nevertheless, because various groups influence the electron transitions to different extents, the error in spectral determination is relatively larger, and so, this red shift varies markedly between individual carotenoid molecules. Recently, an experimental study on the solvent effects of carotenoid spectra was reported in which shorter wavelengths should be expected for transitions in environment with dielectric constant $\epsilon = 1$.⁸⁰

Acknowledgment. This study was supported by grant MSM 0021620835. The Charles University Meta-Centrum in Prague, Czech Republic, is acknowledged for access to its excellent supercomputer facilities. The authors are very grateful to Prof. Tomáš Polívka from South Bohemian University and Profs. Jan Hála and Jakub Pšenčík from our department for very fruitful discussions.

References and Notes

- (1) Blankenship, R. E. *Molecular Mechanisms of Photosynthesis*; Blackwell Science Ltd.: Oxford, U.K., 2002.
- (2) Scheer, H. *Chlorophylls*; CRC Press: Boca Raton, FL, 1991.
- (3) *The Porphyrin Handbook*; Kadish, K. M., Smith, K. M., Guillard, R., Eds.; Academic Press: New York, 2003.
- (4) Gouterman, M. *J. Chem. Phys.* **1959**, *30*, 1139.
- (5) Gouterman, M.; Wagniere, G. H. *J. Mol. Spectrosc.* **1963**, *11*, 108.
- (6) Zerner, M.; Gouterman, M. *Theor. Chim. Acta* **1966**, *4*, 44.
- (7) Almlof, J.; Fischer, T. H.; Gassman, P. G.; Ghosh, A.; Haser, M. *J. Phys. Chem.* **1993**, *97*, 10964.
- (8) Baraldi, I.; Carnevali, A.; Ponterini, G.; Vanossi, D. *J. Mol. Struct. (THEOCHEM)* **1995**, *333*, 121.
- (9) Cortina, H.; Senent, M. L.; Smeyers, Y. G. *J. Phys. Chem. A* **2003**, *107*, 8968.
- (10) Gwaltney, S. R.; Bartlett, R. J. *J. Chem. Phys.* **1998**, *108*, 6790.
- (11) Hasegawa, J.; Ohkawa, K.; Nakatsuji, H. *J. Phys. Chem. B* **1998**, *102*, 10410.
- (12) Hashimoto, T.; Choe, E.; Nakano, H.; Hirao, K. *J. Phys. Chem. A* **1999**, *103*, 1894.
- (13) Kitao, O.; Ushiyama, H.; Miura, N. *J. Chem. Phys.* **1999**, *110*, 2936.
- (14) Lamoén, D.; Parrinello, M. *Chem. Phys. Lett.* **1996**, *248*, 309.
- (15) Merchan, M.; Orti, E.; Roos, B. *Chem. Phys. Lett.* **1994**, *221*, 136.
- (16) Merchan, M.; Orti, E.; Roos, B. *Chem. Phys. Lett.* **1994**, *226*, 27.
- (17) Serrano-Andres, L.; Merchan, M.; Rubio, M.; Roos, B. O. *Chem. Phys. Lett.* **1998**, *295*, 195.
- (18) Nagashima, U.; Takada, T.; Ohno, K. *J. Chem. Phys.* **1986**, *85*, 4524.
- (19) Nakatsuji, H.; Hasegawa, J.-Y.; Hada, M. *J. Chem. Phys.* **1996**, *104*, 2321.
- (20) Parusel, A.; Grime, S. J. *Porphyryns Phthalocyanines* **2001**, *5*, 225.
- (21) Rubio, M.; Roos, B.; Serrano-Andres, L.; Merchan, M. *J. Chem. Phys.* **1999**, *110*, 7202.
- (22) Tokita, Y.; Hasegawa, J.; Nakatsuji, H. *J. Phys. Chem. A* **1998**, *102*, 1843.
- (23) Sundholm, D. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2275.
- (24) Sundholm, D. *Chem. Phys. Lett.* **2000**, *317*, 392.
- (25) Hasegawa, J.; Ozeki, Y.; Ohkawa, K.; Nakatsuji, H. *J. Phys. Chem. B* **1998**, *102*, 1320.
- (26) Parusel, A. B. J.; Grimme, S. *J. Phys. Chem. B* **2000**, *104*, 5395.
- (27) Jeong, D. H.; Jang, S. M.; Hwang, I.-W.; Kim, D.; Yoshida, N.; Osuka, A. *J. Phys. Chem. A* **2002**, *106*, 11054.
- (28) Linnanto, J.; Korppi-Tommola, J. *J. Phys. Chem. A* **2000**, *105*, 3855.
- (29) Linnanto, J.; Korppi-Tommola, J. *Phys. Chem. Chem. Phys.* **2000**, *2*, 4962.
- (30) Linnanto, J.; Oksanen, J. A. I.; Korppi-Tommola, J. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3061.
- (31) Linnanto, J.; Korppi-Tommola, J. E. I. *Phys. Chem. Chem. Phys.* **2002**, *4*, 3453.
- (32) Cory, M. G.; Zerner, M. C.; Hu, X.; Schulten, K. *J. Phys. Chem. B* **1998**, *102*, 7640.
- (33) Sundholm, D. *Chem. Phys. Lett.* **1999**, *302*, 480.
- (34) Sundholm, D. *Phys. Chem. Chem. Phys.* **2003**, *5*, 4265.
- (35) Edwards, L.; Dolphin, D. H.; Gouterman, M. *J. Mol. Spectrosc.* **1970**, *35*, 90.
- (36) Edwards, L.; Dolphin, D. H.; Gouterman, M.; Adler, A. D. *J. Mol. Spectrosc.* **1971**, *38*, 16.
- (37) Houssier, Sauer. *J. Am. Chem. Soc.* **1970**, *92*, 779.
- (38) Konermann, L.; Gatzert, G.; Holzwarth, A. R. *J. Phys. Chem. B* **1997**, *101*, 2933.
- (39) Konermann, L.; Holzwarth, A. R. *Biochemistry* **1996**, *35*, 829.
- (40) Doust, A. B.; Marai, C. N. J.; Harrop, S. J.; Wilk, K. E.; Curmi, P. M. G.; Scholes, G. D. *J. Mol. Biol.* **2004**, *344*, 135.
- (41) McConnell, M. D.; Koop, R.; Vasil'ev, S.; Bruce, D. *Plant Physiol.* **2002**, *130*, 1201.
- (42) Hashimoto, H.; Yoda, T.; Kobayashi, T.; Young, A. J. *J. Mol. Struct.* **2002**, *604*, 125.
- (43) Wang, Y. L.; Mao, L. S.; Hu, X. C. *Biophys. J.* **2004**, *86*, 3097.
- (44) Garavelli, M.; Bernardi, F.; Olivucci, M.; Robb, M. A. *J. Am. Chem. Soc.* **1998**, *120*, 10210.
- (45) He, Z.; Sundstrom, V.; Pullerits, T. *Chem. Phys. Lett.* **2001**, *334*, 159.
- (46) Qian, P.; Mizoguchi, T.; Fujii, R.; Hara, K. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1311.
- (47) Scholes, G. D.; Harcourt, R. D.; Fleming, G. R. *J. Phys. Chem. B* **1997**, *101*, 7302.
- (48) Krueger, B. P.; Scholes, G. D.; Fleming, G. R. *J. Phys. Chem. B* **1998**, *102*, 5378.
- (49) Krueger, B. P.; Scholes, G. D.; Yu, J. Y.; Fleming, G. R. *Acta Phys. Pol. A* **1999**, *95*, 63.
- (50) Krueger, B. P.; Yom, J.; Walla, P. J.; Fleming, G. R. *Chem. Phys. Lett.* **1999**, *310*, 57.
- (51) Scholes, G. D.; Fleming, G. R. *J. Phys. Chem. B* **2000**, *104*, 1854.
- (52) Walla, P. J.; Yom, J.; Krueger, B. P.; Fleming, G. R. *J. Phys. Chem. B* **2000**, *104*, 4799.
- (53) Gradinaru, C. C.; Kennis, J. T. M.; Papagiannakis, E.; van Stokkum, I. H. M.; Cogdell, R. J.; Fleming, G. R.; Niederman, R. A.; van Grondelle, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 2364.
- (54) Hsu, C. P.; Walla, P. J.; Head-Gordon, M.; Fleming, G. R. *J. Phys. Chem. B* **2001**, *105*, 11016.
- (55) Kennis, J. T. M.; Gobets, B.; van Stokkum, I. H. M.; Dekker, J. P.; van Grondelle, R.; Fleming, G. R. *J. Phys. Chem. B* **2001**, *105*, 4485.
- (56) Walla, P. J.; Linden, P. A.; Ohta, K.; Fleming, G. R. *J. Phys. Chem. A* **2002**, *106*, 1909.
- (57) Zimmermann, J.; Linden, P. A.; Vaswani, H. M.; Hiller, R. G.; Fleming, G. R. *J. Phys. Chem. B* **2002**, *106*, 9418.
- (58) Holt, N. E.; Kennis, J. T. M.; Dall'Osto, L.; Bassi, R.; Fleming, G. R. *Chem. Phys. Lett.* **2003**, *379*, 305.
- (59) Ma, Y. Z.; Holt, N. E.; Li, X. P.; Niyogi, K. K.; Fleming, G. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 4377.
- (60) Vaswani, H. M.; Holt, N. E.; Fleming, G. R. *Pure Appl. Chem.* **2005**, *77*, 925.
- (61) You, Z. Q.; Hsu, C. P.; Fleming, G. R. *J. Chem. Phys.* **2006**, *124*, 044506.
- (62) Hsu, C. P.; Head-Gordon, M.; Fleming, G. R. *Abstr. Pap. Am. Chem. Soc.* **2000**, *219*, U329.
- (63) Dreuw, A.; Fleming, G. R.; Head-Gordon, M. *J. Phys. Chem. B* **2003**, *107*, 6500.
- (64) Dreuw, A.; Fleming, G. R.; Head-Gordon, M. *Biochem. Soc. Trans.* **2005**, *33*, 858.
- (65) Hirata, S.; Head-Gordon, M. *Chem. Phys. Lett.* **1999**, *314*, 291.
- (66) Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007.
- (67) Cai, Z.-L.; Crossley, M. J.; Reimers, J. R.; Kobayashi, R.; Amos, R. D. *J. Phys. Chem. B* **2006**, *110*, 15624.
- (68) *Protein Data Bank*; Research Collaboratory for Structural Bioinformatics (RCSB): Piscataway, NJ; <http://www.rcsb.org/pdb/home/home.do>.
- (69) Sundholm, D. *Chem. Phys. Lett.* **2000**, *317*, 545.
- (70) Šeda, J.; Burda, J. V.; Leszczynski, J. *J. Comput. Chem.* **2005**, *26*, 294.
- (71) Burda, J. V.; Šeda, J.; Brázdová, V.; Kapsa, V. *Int. J. Mol. Sci.* **2004**, *5*, 196.
- (72) Chiang, S.-Y.; Lin, I.-F. *J. Chem. Phys.* **2005**, *122*, 094301.
- (73) Chong, D. P.; Takahata, Y. *Chem. Phys. Lett.* **2006**, *418*, 286.
- (74) Starcke, J. H.; Wormit, M.; Schirmer, J.; Dreuw, A. *Chem. Phys.* **2006**, *329*, 39.
- (75) Dreuw, A. *ChemPhysChem* **2006**, *7*, 2259.
- (76) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; Distasio, R. A., Jr.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachselt, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C.-P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L., III; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F., III; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
- (77) Hoff, A. J.; Ames, J. Visible Absorption Spectroscopy of Chlorophylls. In *Chlorophylls*; Scheer, H., Ed.; CRC Press: Boca Raton, FL, 1991; pp 723–738.
- (78) Hála, J.; Šetlík, J. *Biofyzika Fotosyntézy (Scriptum)*; Carolinum Press: Prague, 1999.
- (79) Kohler, E. B. *Carotenoids*; Birkhäuser Verlag: Basel, Switzerland, 1995; Vol. 1B: Spectroscopy.
- (80) Chen, Z.; Lee, C.; Lenzer, T.; Oum, K. *J. Phys. Chem. A* **2006**, *110*, 11291.
- (81) *Phycobiliproteins*; Prozyme: San Leandro, CA, 1999; <http://www.prozyme.com/technical/pbvrwdata.html>.

Dependence of the L-Alanyl-L-Alanine Conformation on Molecular Charge Determined from Ab Initio Computations and NMR Spectra

Vladimír Sychrovský,* Miloš Buděšínský,* Ladislav Benda, Vladimír Špirko,*
Zuzana Vokáčová, Jaroslav Šebestík, and Petr Bour*

Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo nám. 2, 16610, Prague 6, Czech Republic

Received: August 15, 2007; In Final Form: October 21, 2007

The L-alanyl-L-alanine (AA) molecule behaves differently in acidic, neutral, and basic environments. Because of its molecular flexibility and strong interaction with the aqueous environment, its behavior has to be deduced from the NMR spectra indirectly, using statistical methods and comparison with ab initio predictions of geometric and spectral parameters. In this study, chemical shifts and indirect spin–spin coupling constants of the AA cation, anion, and zwitterion were measured and compared to values obtained by density functional computations for various conformers of the dipeptide. The accuracy and sensitivity of the quantum methods to the molecular charge was also tested on the (mono)-alanine molecule. Probable AA conformers could be identified at two-dimensional potential energy surfaces and verified by the comparison of the computed parameters with measured NMR data. The results indicate that, whereas the main-chain peptide conformations of the cationic (AA⁺) and zwitterionic (AA^{zw}) forms are similar, the anion (AA[−]) adopts also another, approximately equally populated conformer in the aqueous solution. Additionally, the NH₂ group can rotate in the two main chain conformations of the anionic form AA[−]. According to a vibrational quantum analysis of the two-dimensional energy surfaces, higher-energy conformers might exist for all three charged AA forms but cannot be detected directly by NMR spectroscopy because of their small populations and short lifetimes. In accord with previous studies, the NMR parameters, particularly the indirect nuclear spin–spin coupling constants, often provided an excellent probe of a local conformation. Generalization to peptides and proteins, however, has to take into account the environment, molecular charge, and flexibility of the peptide chain.

Introduction

NMR spectroscopy has a long history in the conformational analyses of peptide structures.^{1,2} Empirical correlations of chemical shifts and nuclear spin–spin coupling constants (*J*-coupling) with the geometry were originally used to discriminate helical and sheetlike peptides and subsequently were extended to protein studies.^{3–5} The possibility to calculate the NMR parameters for larger molecules with reasonable precision has lately provided an additional basis for the interpretation of the experiment, which consequently facilitated the verification of various conformational models. Particularly, the analytical approaches to chemical shifts⁶ and the coupling constants^{7–9} within the density functional theory (DFT) speeded up the computations and facilitated the conformational studies of interesting peptide systems.^{10,11} The computations improved, for example, the empirical Karplus relations between the spectra and the structure.^{12,13} However, peptide flexibility, solvent effects, and local vibrational motions have to be taken into account in accurate modeling.¹⁴

The dependence of the L-alanyl-L-alanine (AA) conformation on molecular charge studied in this work thus provides additional information about the behavior of the peptide chain in various pH conditions as well as about the accuracy and validity of current simulation techniques. NMR properties of charged and zwitterionic peptides themselves are notoriously difficult to

calculate¹⁵ mainly because of the electronic charge concentration requiring a large basis set and because of the strong interaction with the environment, in most cases with water.¹⁶ The three AA forms also provide an experimentally well-accessible example of a simple molecular mechanical system controllable by pH. Therefore, we find it interesting to analyze in detail the two-dimensional potential energy surface and account for possible vibrational quantum effects.¹⁷

In the second part of this work, we use the statistical comparison of the experimental and computed chemical shifts and spin–spin coupling constants developed previously for the AA zwitterion.¹¹ The ability of the computation to discriminate between various charged forms is tested on the alanine molecule (A) labeled with stable ¹⁵N and ¹³C isotopes, where the conformational problem is simpler than in AA. The influence of the charge on the molecular potential energy surface, expressed as a function of the main-chain peptide torsion angles (φ , ψ), is computed with the inclusion of a continuum solvent correction. It appears that the pH (charge) change stabilizes a new anion (AA[−]) conformer and that the conformational equilibrium can be proven from the NMR data. The quantum vibrational analysis predicts also other well-defined conformers, which, however, are neither significantly populated under normal conditions nor can be observed directly by NMR due to their short lifetimes.

Experimental Section

Isotopically labeled L-alanine (¹³C,98%; ¹⁵N,98%) was purchased from Stable Isotopes, Inc., whereas the nonlabeled AA

* To whom correspondence should be addressed: E-mail: vladimir.sychrovsky@uochb.cas.cz (V.S.), budes@uochb.cas.cz (M.B.), vladimir.spirko@marge.uochb.cas.cz (V.S.), bour@uochb.cas.cz (P.B.).

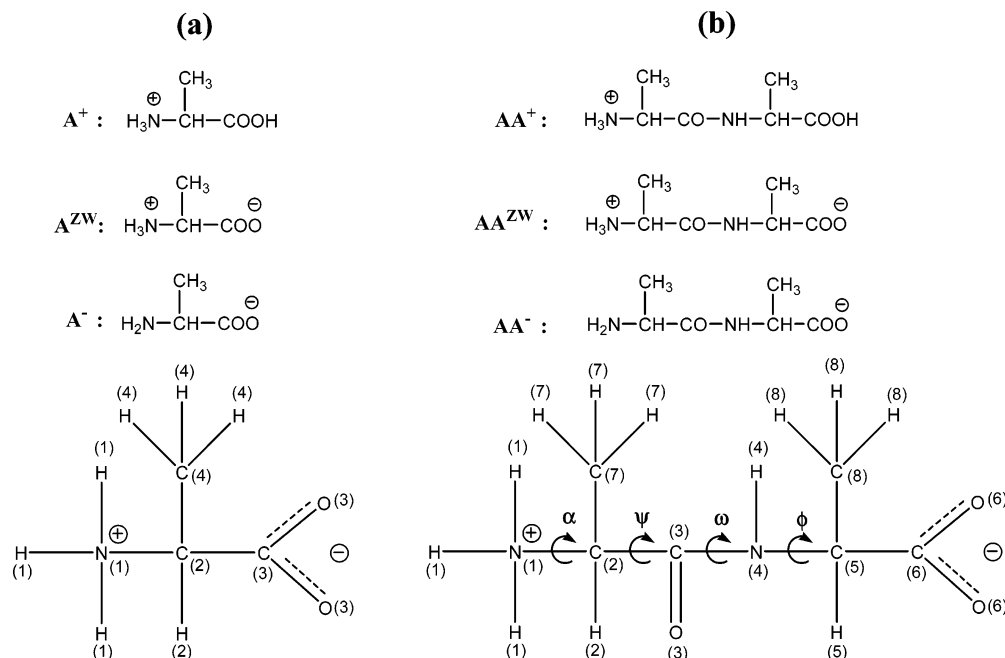


Figure 1. The ionic forms and symbols used for L-alanine (a) and AA (b). The numbering of the atoms for the definition of NMR parameters is shown in zwitterionic forms of both molecules. In addition to standard peptide torsion angles (ϕ , ψ , ω), we have introduced the angle α as the average angle of the two amine hydrogen $-(C3-C2-N1-H1)$ torsion angles (for the NH_2 residue in AA^-).

was purchased from Sigma. The synthesis of labeled AA is described elsewhere.¹¹ NMR spectra of labeled alanine (A^+ , A^- , A^{ZW}), natural AA (AA^+ , AA^- , AA^{ZW}), and labeled AA^{ZW} were measured with Fourier transform (FT) NMR spectrometers Varian UNITY-500 and Bruker AVANCE-500 (1H at 500 MHz, ^{13}C at 125.7 MHz, ^{15}N at 50.7 MHz, ^{17}O at 67.8 MHz) in D_2O and/or in the mixture H_2O/D_2O (9:1). The solution pH was varied by additions of 2 M HCl (pH \approx 2) and NaOH (pH \approx 12) solutions. For these pH values, the AA peptide exists exclusively in the AA^+ and AA^- forms, respectively, which was confirmed by measuring of the NMR titration curves (provided in the Supporting Information (SI), together with relevant pK constants for A and AA). The zwitterionic forms were obtained by dissolving the compounds in distilled water without any buffer. All spectra were measured at room temperature. Chemical shifts were referenced either to internal (DSS for 1H and ^{13}C and H_2O for ^{17}O , but the oxygen is not discussed in this work) or external (nitromethane in the capillary for ^{15}N) standards. The structural assignment of the hydrogen and carbon chemical shifts was achieved using homonuclear and heteronuclear two-dimensional techniques with pulse field gradients (2D- 1H , 1H -PFG-COSY, 1H , ^{13}C -PFG-HSQC, and 2D- 1H , ^{13}C -PFG-HMBC) in D_2O . The solvent mixture H_2O/D_2O (9:1) was used to observe the signals of NH and NH_3^+ protons. Only the couplings of amide NH could be observed in this mixture because of the fast exchange rate of amine NH_3^+ protons with water. The $J(H,H)$ values were determined from the 1D- 1H NMR spectrum and the $J(C,H)$ couplings from the non-decoupled 1D- ^{13}C NMR spectrum. A series of selective 1H -decoupled ^{13}C NMR spectra was used to assign individual $J(C,H)$ couplings. The labeled ^{15}N and ^{15}N , ^{13}C AA samples were used mainly in order to obtain $J(N,H)$, $J(N,C)$, and $J(C,C)$ coupling constants using the 1D- 1H and ^{13}C NMR spectra, the 1D- ^{13}C -INADEQUATE (Incredible Natural Abundance Double Quantum Transfer Experiment), and the 2D- 1H , ^{15}N -PFG-HMBC spectra.

Calculations. The GAUSSIAN software¹⁸ was used for the quantum chemical computations. The torsion angles ϕ and ψ (Figure 1) were varied with 30° steps, and for each of the

resultant $12 \times 12 = 144$ geometries, all the remaining coordinates were fully relaxed by energy minimization. The angle ω was initially set to 180° so as to maintain the *trans*-peptide bond, because the experimental data do not suggest a presence of the *cis*-conformer. In the scans, the relaxed ω angle deviated from 180° by less than $\sim 5^\circ$. The BPW91¹⁹ functional and the standard Pople-type 6-311++G** basis were used with the PCM solvent model for water.²⁰ For the anion AA^- , three scans were performed (3×144 points), taking into account three initial orientations ($\alpha = -120, 0$, and 120°) of the NH_2 group. For this purpose, we defined the angle α as an average of the two amine hydrogen $-(C3-C2-N1-H)$ angles. The geometries of the local minima located on the resultant surfaces were fully optimized without any constraints. The scan of the zwitterion mimics previous computations done with smaller grid steps.¹¹ For control computations, other potential energy scans were done in a vacuum for AA^+ and AA^- (AA^{ZW} is not stable without solvent), and local minima geometries were reoptimized at the MP2²¹/6-311++G** level of approximation.

For the grid points and local minima geometries, the NMR shielding tensors and the indirect NMR spin-spin coupling constants were calculated. The default gauge-invariant atomic orbital (GIAO)²² method was used for the shieldings to prevent their origin dependence. For the J -couplings, all four important terms^{7,23} were included in the analytical coupled-perturbed DFT computation. For the shieldings and J -couplings, we used the B3LYP functional²⁴ with the IGLOII or IGLOIII bases,²⁵ which are believed to be well-suited for computations of the NMR properties.⁸ In all cases, the same PCM solvent model was applied. With the same method, the NMR parameters were calculated for the anion, cation, and zwitterion of (mono)-alanine in equilibrium geometries. Chemical shifts were related to standard molecules (DSS for 1H and ^{13}C , nitromethane for ^{15}N , water for ^{17}O , same as in ref 11).

The quantum and dynamical effects of the torsional motions involving the angles (ϕ , ψ) were studied using an approximate Hamiltonian described in full in ref 11, where the potential was obtained by fitting of the computed two-dimensional surfaces, and the kinetic part was expressed in the curvilinear angular

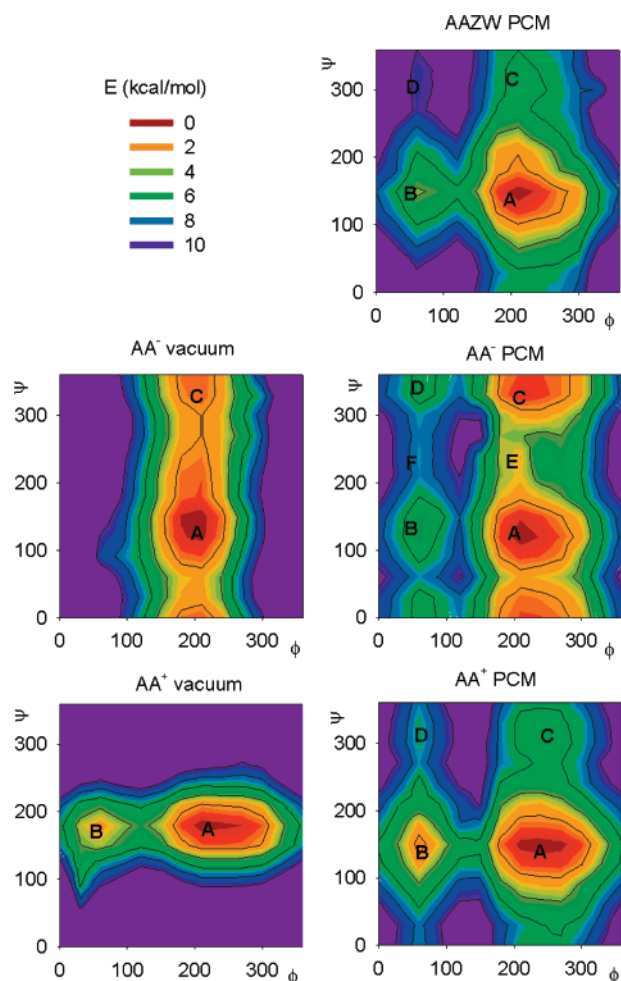


Figure 2. Contour plots of the computed (BPW91/6-311++G**) potential energy surfaces of the three AA forms. By default, the PCM solvent model was applied (right-hand side); for the AA⁺ and AA[−] forms stable in a vacuum, the dependencies without the solvent are shown on the left. Approximate positions of local minima are marked by the capital letters. To avoid splitting of the minimum well, the angles are plotted within the positive (0, 360°) interval instead of the usual (−180, 180°) range.

coordinates.²⁶ The eigenvalue vibrational problem $H\Psi = E\Psi$ was solved variationally in basis set functions expressed as products of the eigenfunctions of the corresponding uncoupled one-dimensional Schrödinger equations. The one-dimensional functions were determined numerically using the Numerov–

Cooley integration procedure²⁷ and used for averaging the NMR parameters. The averaging, however, did not bring significant corrections with respect to the overall accuracy and is not discussed further. The lifetimes of selected localized excited vibrational states were estimated from simplified one-dimensional modeling.^{11,28}

Results and Discussion

AA Conformers. The calculated adiabatic energy dependencies on the φ and ψ torsion angles are presented in Figure 2. Apparently, the energy profiles of the three hydrated forms exhibit several similarities. The global minimum **A** is associated with comparable angle values for all cases (see Table 1 for details). The similarity of the AA^{ZW} and AA⁺ surfaces is the most obvious. However, the protonation of AA^{ZW} makes the resultant cation (AA⁺) more flexible with respect to the φ rotation, and the potential well is elongated along this coordinate. The MP2 method provides an equilibrium value of the φ angle (−151°, see Table 1) even significantly shifted from the DFT angle of −121°. For AA⁺, the local energy minimum **B** deepens when compared to the other forms, but its relative energy of 1.4 kcal/mol (cf. Table 1) and narrow potential well probably still prevent a significant population of this conformer in the sample. Obviously, the least probable is an occurrence of the other two, **C** and **D** conformers of AA^{ZW} and AA⁺.

The energy surface of the anion (AA[−], the middle of Figure 2) is different. Whereas the geometry of the lowest-energy **A** conformer is very close to that of the zwitterion, the minimum well **C** significantly broadens, and its energy is comparable with the global minimum **A**. Additionally, new, very shallow minima (**E**, **F**) appear for the anion; these are, along with the **B** and **D** extremes, not populated due to their high relative energies. For **C**, however, the computed relative energy (0.4 kcal/mol, Table 1) is probably comparable with the computational error and suggests a significant presence of this conformer in the sample at room temperature.

The more complicated behavior of the anion AA[−] stems predominantly from the directional and ambivalent binding properties of the NH₂ group. The nitrogen electron lone pair can make an internal hydrogen bond to the amide hydrogen, or the NH₂ protons can be bonded to the carbonyl oxygen. This is documented in Figure 3, where for a fixed value of $\varphi = -150^\circ$ the energy dependence on the ψ angle is plotted. Although detailed three- or more dimensional energy scans are currently not feasible, the coupling of the α , ψ , and φ torsional motions

TABLE 1: Computed (BPW91/PCM/6-311++G) Geometries and Relative Conformer Energies of the Three Charged AA Forms^a**

conformer	AA ⁺			AA ^{ZW}			AA [−]			
	ψ	φ	E	ψ	φ	E	α	ψ	φ	E
A	149	−121	0.0	147	−153	0.0	−2	127	−152	0.0
A'							118	128	−151	0.8
A''							−125	119	−150	1.9
A ^b	178	−159					−11.3	138	−160	
A ^c	150	−151		146	−158		−13.1	133	−157	
A ^d	169	−160					−2.5	139	−160	
B ^e	150	60	1.4	150	66	3.6	−3	135	63	4.3
C	−55	−127	5.1	−51	−150	5.1	−136	−19	−152	0.4
C'							145	8	−147	0.8
C''							9	−21	−150	1.4
D	−55	53	6.3	−48	64	9.2	−138	−16	63	5.0
E							5	−105	−154	2.7
F							7	−116	61	6.9

^a Torsion angles (φ , ψ , α) are given in degrees and the relative energies (E) in kcal/mol. ^b BPW91, gas phase. ^c MP2, PCM. ^d MP2, gas phase. ^e Calculated lifetimes of this conformer are approximately 1, 6–10, and 0.7 ms for AA⁺, AA^{ZW}, and AA[−], respectively.

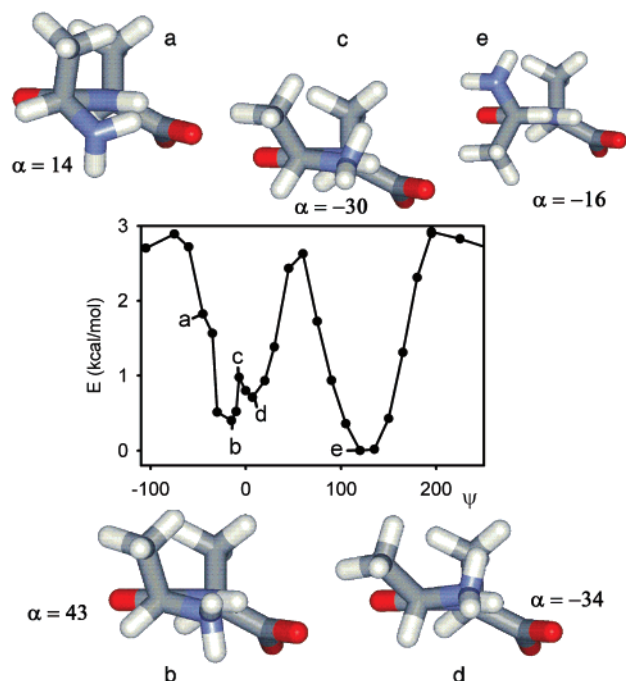


Figure 3. Detailed one-dimensional profile of the AA^- calculated potential energy surface for $\varphi = -150^\circ$. The NH_2 binding pattern changes along the lowest-energy path: (a) the NH_2 hydrogens may be attracted by the COO^- group and the amide bond π -system, or (b) the NH_2 nitrogen electron lone pair creates a hydrogen bond to the amide hydrogen which is (c, d) conserved during the rotation over the syn ($\psi \approx 0^\circ$) conformation. At the other minimum (e), the NH_2 hydrogens are attracted to the π -system as well as to the amide oxygen. The NH_2 rotation is indicated (α , in degrees).

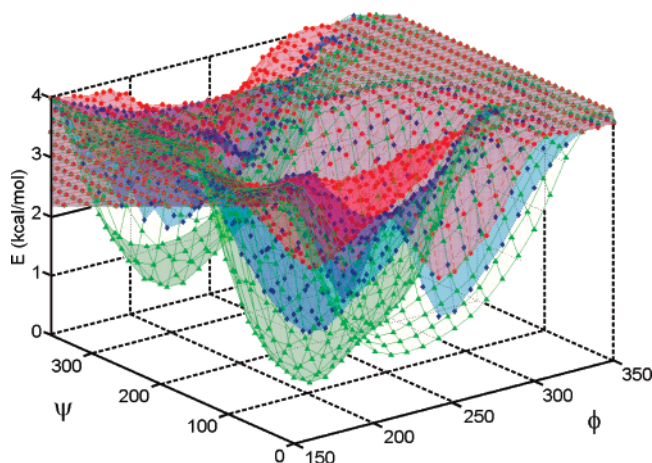


Figure 4. Part of the anionic (AA^-) potential energy surfaces (green, blue, and red) calculated for three NH_2 rotamers.

in AA^- nicely reveals the complexity and caveats that must be taken into account in the modeling of peptide conformational landscape. In longer molecules and proteins the folding of the backbone can presumably be influenced by side-chain interactions similarly as the (ψ, φ) energy map by the NH_2 group rotation. The one- and two-dimensional surfaces (Figures 2 and 3) can be also considered as projections of multidimensional surfaces. Another way illustrating this situation is represented in Figure 4, where three (φ, ψ) anionic surfaces are plotted as obtained from scans with three different NH_2 rotamers. The lowest-energy surface in Figure 2 was obviously obtained by taking the minimum energies for each (φ, ψ) pair in Figure 4.

However minor the purely electrostatic influence of the charged ends on the conformation might be in comparison with

the role of the NH_2 group, a more detailed look at the optimal geometries in Table 1 reveals effects that can be attributed solely to the charge/pH changes. The protonation of the COO^- group ($AA^{ZW} \rightarrow AA^+$) results in a change of the φ torsion around the adjacent C–N bond by about 30° with the ψ torsion remaining nearly unaffected. Similarly, for the anionic form the ψ torsion decreases by about 20° under the $NH_3^+ \rightarrow NH_2$ deprotonation of the zwitterion. Whereas the effects of the pH and the NH_2 intrinsic hydrogen bonding can hardly be separated, the charge change undoubtedly further tweaks the dipeptide conformational properties.

Strictly speaking, two rotamers of the $COOH$ group (and two other rotamers associated with the OH rotation are theoretically possible) should be considered in the potential energy surface of the cation AA^+ in a similar way as the NH_2 rotation was treated in the anion AA^- . However, the energy changes involved in the $COOH$ rotation are minor (less than 1 kcal/mol) and so is the effect of these geometry variations on the NMR parameters. Therefore, the AA^+ (φ, ψ) surface was considered only for the lowest-energy $COOH$ conformers.

Not all local minima on the potential energy surfaces in Figure 2 can support a stable quantum state. For both AA^+ and AA^- , only the two lowest-energy conformers are stable. This can be deduced from the two-dimensional Hamiltonian and the resultant localized wave functions shown for the cation and anion in Figure 5. A similar result was obtained for the AA zwitterion.¹¹ The transition barriers are clearly high enough to support at least the two distinct conformers with the wavefunctions plotted by the green lines in Figure 5. Moreover, we can deduce from the lifetimes of the second-lowest energy states listed in the footnotes of Table 1 (a few milliseconds) that it is in principle possible to prepare these conformers, although presently they cannot be detected by the inherently slow NMR experiment.

By comparison of the computations performed in a vacuum and with the solvent correction (the left- and right-hand parts of Figure 2, respectively), one can well estimate the effect of the environment on the peptide conformational properties. While it is true that the basic conformational characteristics of the dipeptide given by the covalent bonds do not change upon the solvation, the final energy profiles, the exact minima geometries, and, particularly, the steepness of the equilibrium potential wells do change. The MP2 method provides virtually the same conformers as the BPW91 functional (Table 1) except for the φ angle of the AA^+ form; in this case, however, the global minimum well is very broad and the equilibrium φ difference does not necessarily implicate a different behavior. Other backbone torsion angles obtained for the minima with the BPW91 and MP2 methods differ by less than 6° . The values of the torsion angles for the A-type AA conformers are close to those observed in β -sheet structures of longer peptides and proteins (antiparallel β -sheet: $\varphi, \psi = -139^\circ, 135^\circ$; parallel β -sheets: $\varphi, \psi = -119^\circ, 113^\circ$),²⁹ whereas the other types do not have canonical protein counterparts.

Similarly as for the zwitterion,¹¹ molecular dynamics (MD) simulations provided analogous conformer distribution to those deduced from the ab initio relative energies also for the positively and negatively charged AA forms (see Figure 1s in SI for the Amber force field). Particularly, the MD computations are consistent with the one (for AA^{ZW} and AA^+) and two (for AA^-) conformer presence in the sample under room temperature. Obviously, detailed MD results strongly depend on the force field parametrization, and presently we consider them inferior to those based on the DFT energy maps.

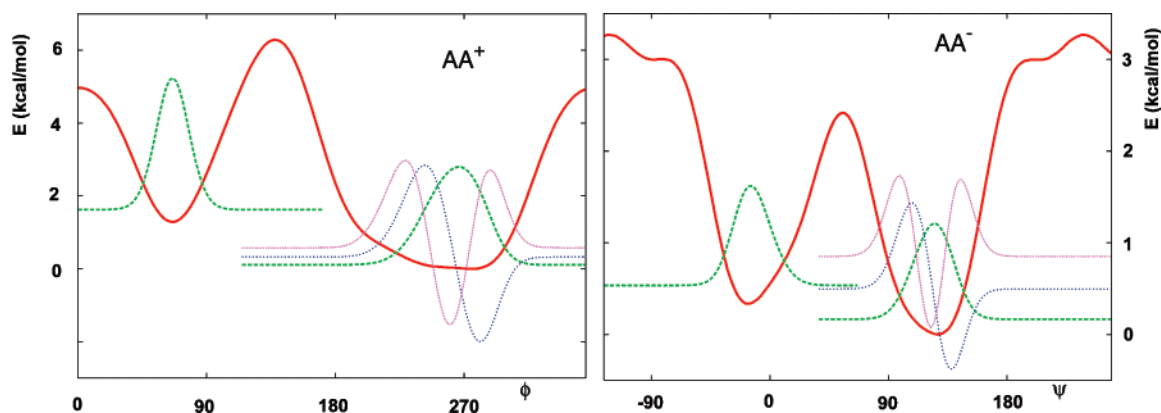


Figure 5. One-dimensional sections of the cation (AA^+) and anion (AA^-) smoothed potential energy surfaces (the red line) and vibrational wave functions of the lowest-energy states (the dashed/dotted lines, with the asymptotes corresponding to their energies).

TABLE 2: Experimental (in D_2O) and Computed^a Changes of Alanine NMR Parameters under the pH Variations

	$\Delta(A^+ - A^{ZW})$		A^{ZW}		$\Delta(A^- - A^{ZW})$	
	calcd	exptl	calcd	exptl	calcd	exptl
Chemical Shifts (ppm)						
N	-4.00	-2.20	-386.1	-339.6	-13.1	-6.6
C2	-0.17	-1.77	62.79	53.35	1.32	0.93
C3	-0.26	-3.11	188.38	178.66	11.74	8.94
C4	-0.27	-0.83	20.70	18.97	8.14	4.25
H2	0.81	0.37	4.17	3.78	-0.64	-0.48
H4	0.18	0.08	1.57	1.48	-0.44	-0.26
Spin-Spin Coupling Constants (Hz)						
¹ J:						
N1,C2	-1.6	-0.85	-3.2	-5.7	1.9	1.4
C2,C4	-4.1	-0.8	36.8	34.9	-1.9	0.3
C2,C3	12.2	5.6	48.5	54.0	3.1	-1.3
C2,H2	4.4	1.5	143.6	145.1	-10.4	-6.7
C4,H4	2.9	1.3	127.0	129.7	-3.4	-2.1
² J:						
N1,H2	-0.2	0	-0.9	0	-2.4	-2.2
N1,C4	-0.1	0	-0.7	0	-3.1	0
N1,C3	0.1	0	-0.5	0	0	0
C2,H4	0.2	-0.2	-3.1	-4.4	0.3	0.1
C4,H2	-0.8	-0.35	-3.1	-4.55	1.0	-0.15
C3,C4	-0.3	-0.1	-0.7	-1.2	0.4	1.2
C3,H2	-0.8	-1.0	-4.6	-5.0	0.5	0.7
³ J:						
N1,H4	0.3	0.05	-3.3	-3.05	0.6	0.1
C3,H4	0.5	0.35	4.1	4.2	-0.2	0.1
H2,H4	0.1	0	7.3	7.3	-0.6	-0.2

^a BPW91/6-311++G**/PCM geometries, B3LYP/IGLOIII NMR parameters. For A^+ and A^- , differences with respect to A^{ZW} are given; for A^- , the values calculated for three NH_2 rotamers were averaged.

Alanine NMR pH Dependence. The anion (A^-), cation (A^+), and zwitterion (A^{ZW}) of the ^{15}N , ^{13}C -isotopically labeled alanine allowed us to investigate the dependence of the chemical shifts and spin-spin coupling constants for a system where no significant change of molecular shape can be induced by the change of the molecular charge. A complete set of experimental and calculated NMR parameters is given in Table 2. The absolute chemical shifts are reproduced with a notable error (e.g., computed 53.35 ppm, experimental 62.79 ppm for C2); nevertheless all signs of the differences between the charged and neutral forms are reproduced correctly. As expected, the hydrogen shifts could be calculated with a higher precision than those for the heavy atoms.

The alanine spin-spin coupling constants change under the deprotonation and protonation like the shifts, typically within 1–10%. The computation well reproduces the magnitudes of the individual coupling constants (e.g., for A^{ZW} , the experimental

$^1J(C2,H2) = 145.1$ Hz was calculated as 143.6 Hz, i.e., with 1% error). As in the case of the shifts, the computations reproduce most trends induced by the pH change, albeit with a limited precision. As an extreme case, the $^1J(C2,C3)$ coupling changes by -1.3 Hz while a +3.1 Hz change is predicted by the theory. With the exception of the error of the DFT method, we attribute the deviations to MD or solvent-solute interactions, incompletely covered by the present model. Other functionals (BPW91) and bases (IGLOII, 6-311++G**) provided similar results. Overall, we can see that the computations correctly reproduce the main changes in NMR shifts and coupling patterns induced by the change of molecular charge.

Accuracy of the Calculated NMR Parameters in AA. For the AA^+ cation, the accuracy of various approximate levels used for the computations of the NMR shifts and coupling constants are demonstrated in Table 3. The anion AA^- behaves similarly; the data can be found in the SI (Table 1s). As for alanine, the error of the chemical shifts computed for dialanine significantly exceeds the estimated experimental inaccuracy (estimated as 0.01, 0.02, and 4 ppm for the NMR shifts of hydrogen, carbon, and nitrogen, respectively). As observed earlier,^{11,30} the hydrogen shifts are reproduced relatively accurately, whereas the DFT method becomes inaccurate for carbons, with the inaccuracy being even higher for the nitrogen atom. The parameters computed at different approximation levels do not vary dramatically. The best overall agreement of the calculated data with the experiment was achieved with the BPW91/6-311++G**/PCM equilibrium geometry and at the B3LYP/IGLOII/PCM level for the NMR parameters. The IGLOIII basis, although bigger, provides less accurate shifts than IGLOII.

Interestingly enough, the gas-phase computations appear to be reasonably accurate. For the geometry obtained with the MP2 method in the gas phase (vacuum), for example, the calculated carbonyl carbon shifts are even closer to experiment than when the PCM solvent correction is applied. For DFT, however, the PCM results are more precise. The differences in geometries (Table 1) and NMR shifts obtained with the MP2 and DFT and with PCM and gas phase might indicate that the current potential energy surfaces (Figure 2) are not quite accurate and that a better solvent model accounting for the directional hydrogen bonds³¹ would be more appropriate; this is, however, impossible to achieve with the computer means available. On the other hand, some properties of the dipeptide are reproduced very well, e.g., the observed 84 ppm shift difference between N1 and N4 was calculated within the 92–95 ppm interval. The chemical environment (amine-amide) is thus perhaps more reproducible than solvent environment (vacuum-water), whose influence is weaker.

TABLE 3: Chemical Shifts and Spin–Spin Coupling Constants Calculated at Different Levels of Theory for Conformer A of the AA⁺ Cation^a

geometry (6-311++G**): NMR (B3LYP):	BPW91 PCM IGLOIII/PCM	BPW91 PCM IGLOII/PCM	BPW91 (gas) IGLOIII (gas)	MP2 PCM IGLOIII/PCM	MP2 (gas) IGLOIII (gas)	exptl
Chemical Shifts (ppm)						
H2	4.63	4.45	4.07	4.40	3.87	4.10
H5	5.53	5.27	4.75	5.01	4.58	4.42
H7	1.58	1.58	1.77	1.41	1.56	1.55
H8	1.48	1.45	1.69	1.31	1.52	1.45
C2	63.2	61.4	63.3	59.9	60.0	51.8
C3	182.6	177.4	178.7	181.5	176.0	173.5
C5	56.4	54.6	61.8	55.7	58.9	51.6
C6	192.2	186.6	187.9	189.8	185.7	179.1
C7	22.0	21.5	21.0	21.8	22.2	19.4
C8	24.3	23.8	21.8	23.2	21.4	18.9
N1	−390	−379	−390	−392	−392	−342
N4	−295	−288	−297	−300	−299	−258
$\Delta\sigma$	11.09	8.27	10.67	10.89	10.11	
Coupling Constants (Hz)						
¹ J:						
C7,H7	129.1	129.8	130.2	128.6	129.9	130.7
C8,H8	128.7	129.3	130.7	128.1	129.8	130.4
C5,H5	140.1	141.8	144.6	142.1	144.1	146.8
C2,C3	52.1	54.5	45.4	51.9	46.1	52.4
C5,C6	60.6	63.7	58.7	60.1	58.6	58.9
C2,C7	32.7	34.6	33.7	32.0	31.8	33.7
C5,C8	33.7	35.6	32.6	32.2	32.1	34.4
C2,N1	−4.6	−5.0	−6.0	−0.2	−0.6	−9.2
C3,N4	−18.4	−19.4	−23.0	−18.4	−22.5	−17.0
C5,N4	−11.8	−12.6	−10.8	−12.3	−11.3	−11.5
² J:						
C2,H7	−2.6	−3.2	−2.6	−2.8	−2.6	−4.4
C5,H8	−2.9	−3.5	−2.9	−3.0	−3.0	−4.5
C7,H2	−2.6	−3.0	−2.8	−2.5	−2.4	−3.9
C8,H5	−4.7	−5.0	−3.8	−4.1	−3.7	−4.2
C6,H5	−6.2	−6.9	−5.7	−6.3	−5.7	−4.2
C6,C8	0.2	0.1	−1.2	−0.6	−1.3	1.1
C7,N1	−0.5	−0.5	−0.4	−0.5	−0.5	1.3
C2,N4	−11.1	−11.6	−8.5	−11.0	−9.2	−6.9
C6,N4	−1.0	−1.0	−1.1	−1.2	−1.0	−1.1
³ J:						
H2,H7	7.0	6.7	7.2	6.9	7.2	7.1
H5,H8	7.3	7.1	7.0	7.2	7.1	7.0
C3,H7	4.3	4.4	4.6	4.2	4.3	4.4
C6,H8	4.4	4.5	4.7	4.4	4.6	4.5
N1,H7	−3.8	−3.8	−3.7	−3.8	−4.0	−3.1
N4,H8	−2.8	−2.8	−2.9	−2.5	−2.9	−3.1
C2,C5	2.4	2.5	2.2	2.5	2.2	2.1
ΔJ	1.4	1.5	1.4	1.6	1.7	

^a $\Delta\sigma$ and ΔJ are average absolute deviations. Isotropic shielding values of (31.83, 183.45, and −180.16 ppm, for the H, C, and N atoms, respectively) and (31.53, 181.10, and −194.36 ppm) were used for the IGLOII and IGLOIII computations.

The scalar couplings calculated using different approaches (the lower part of Table 3) also vary rather moderately. Like the shifts, the average absolute deviations ranging in a narrow interval of 1.4–1.7 Hz do not favor any particular method. No preferential approach was indicated even by a decomposition of the statistics into the absolute average deviations for the ¹J, ²J, and ³J couplings (not shown). The final precision seems to be an internal property of the B3LYP functional and perhaps the DFT methodology;³² it was also discussed in previous works.^{11,33} An occasional generalization, however, can be made. Particularly, the application of the PCM method improves some one-bond couplings (¹J(C2,C3) and ¹J(C3,N4)) for both the DFT and MP2 geometries. On the contrary, some couplings deviate more from the experiment upon the application of PCM (²J(C2,N4) and ¹J(C5,H5)). As expected, the charged and polar groups are the most sensitive to the PCM–vacuum environment change.

Behavior similar to that of the calculated cation chemical shifts was also observed for the anion (Table 1s in the SI),

although here the situation is complicated by the influence of the NH₂ group rotation. Indeed, a relatively large dispersion of the calculated shifts appeared for the nuclei in the vicinity of the amine group. For the three NH₂ rotamers of the conformer A, the calculated shifts are, for example, 24.7, 24.7, and 29.5 ppm for carbon C7, 63.3, 63.9, and 63.7 ppm for carbon C2, and 193.8, 190.7, and 190.6 ppm for carbon C3. The dispersion is even bigger for the conformer C: 27.1, 25.9, and 25.7 ppm for carbon C7, 61.9, 62.9, and 65.5 ppm for carbon C2, and 190.4, 191.5, and 193.7 ppm for carbon C3.

AA Chemical Shifts. Similarly as for the alanine, the computed and experimental chemical shift changes for the charged AA forms (related to AA^{ZW}, Figure 6) confirm that the theory can reproduce the experiment on average but with a limited accuracy. Extreme changes are usually better reproduced than the small ones. Especially the hydrogen shift changes are smaller than 0.5 ppm, with the exception of H1 (experimental values are σ = 6.16 ppm for AA^{ZW} and 8.09 ppm for AA⁺)

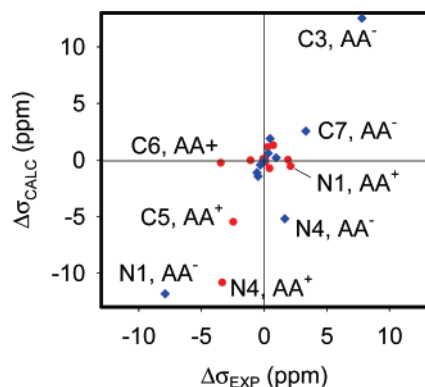


Figure 6. Comparison of calculated and experimental pH chemical shift changes (for the lowest-energy AA^+ and AA^- conformers, with respect to the zwitterionic form AA^{ZW}). For the biggest changes, the corresponding atoms are indicated in the plot.

and not easily reproducible. The amide NH group is also problematic: the computation overestimates the observed nitrogen (N4) shift both in the anion and cation. On the other hand, the amide carbon (C3) change in AA^- is easily reproduced; it is clearly caused by the loss of charge at the amine group. Large charge-induced shift variations (exceeding 1 ppm) have also been measured for the C5, C6, and N1 AA^+ atoms, all indicated in Figure 6. The calculated carbon shifts (conformer **A** in cation, zwitterion, and anion) of C3 (182.6, 181.3, 193.8 ppm), C5 (56.4, 61.9, 60.5 ppm), C6 (192.2, 192.5, 194.4 ppm), and C7 (19.4, 19.3, 22.6) thus nicely correspond to the trend observed under the pH change experimentally: C3 (173.5, 172.8, 180.6 ppm), C5 (51.6, 54.1, 53.6 ppm), C6 (179.1, 182.5, 183 ppm), and C7 (22.0, 22.2, 24.7 ppm). This reflects a general tendency of the *ab initio*/DFT computation to reproduce relative values of NMR parameters with a higher accuracy than for absolute ones.³⁴

The accuracy of the calculated shifts (Table 2s in Supporting Information), however, is not sufficient to discriminate between the individual AA conformers **A–F**. For example, the absolute overall deviations from the experiment range within 11.1–11.7, 11.8–16.5, and 11.3–11.9 ppm (for the cation, zwitterion, and anion, respectively) only. All three AA forms thus behave similarly. The variations of the calculated NMR shifts should

then be attributed to several factors, such as the reaction of the PCM continuum to the charge redistribution in the charged peptide forms, and do not directly reflect detailed conformational changes.

AA Spin–Spin Coupling. As documented in Figure 7, the coupling constant pH variation can be reproduced with the computation similarly as the shifts. Also here, smaller changes are less reliably calculated than the bigger ones, and the theoretical values concerning the vicinity of the charged residues are less accurate. The change of $^1J(\text{C5}, \text{H5})$ in AA^+ , for example, was predicted at the opposite direction. The magnitudes of the one-bond (1J) constants change most, but the biggest relative changes can be found between the vicinal and geminal couplings ($^2J, ^3J$).

Unlike for the shifts, calculated coupling constants provide useful information on AA conformation. This can be seen for the anion AA^- in Table 4, where the measurable coupling constants calculated for 10 conformers are listed and compared to the experiment. We can see that the results are consistent with the estimated relative conformer energies: The coupling constants calculated for the energetically inconvenient conformers (**B**, **D–F**) significantly deviate from the experimental values, which is also indicated by the average absolute deviations from the experiment listed at the bottom of the Table. The conformer **C** alone exhibits the lowest average deviations in the constants, whereas the **A** conformer is more preferred energetically. However, this can be explained by the conformational equilibrium. Indeed, the Boltzmann averaging, taking into account all the conformers **A–F** including the NH_2 group rotamers, provides a reasonably low average deviation of the couplings. Obviously, as discussed above and observed in previous works,^{11,33} a future improvement of the computational model is desirable because more accurate theoretical constants can lead to a better discrimination between the peptide conformers.

For example, the NMR spectra would be sensitive to the NH_2 group rotation as the constants computed for the three rotamers differ significantly (for conformer **A**, the calculated $^1J(\text{C2}, \text{H2})$ are 135.9, 134.0, and 136.8 Hz, $^1J(\text{C2}, \text{C3})$ 48.3, 53.4, and 53.4 Hz, $^2J(\text{N1}, \text{H2})$ –3.3, 0.7, and –3.9 Hz, etc.). The NH_2 rotation, however, influences the couplings only locally with remote atomic groups not being affected: The constants $^1J(\text{N4}, \text{H4})$

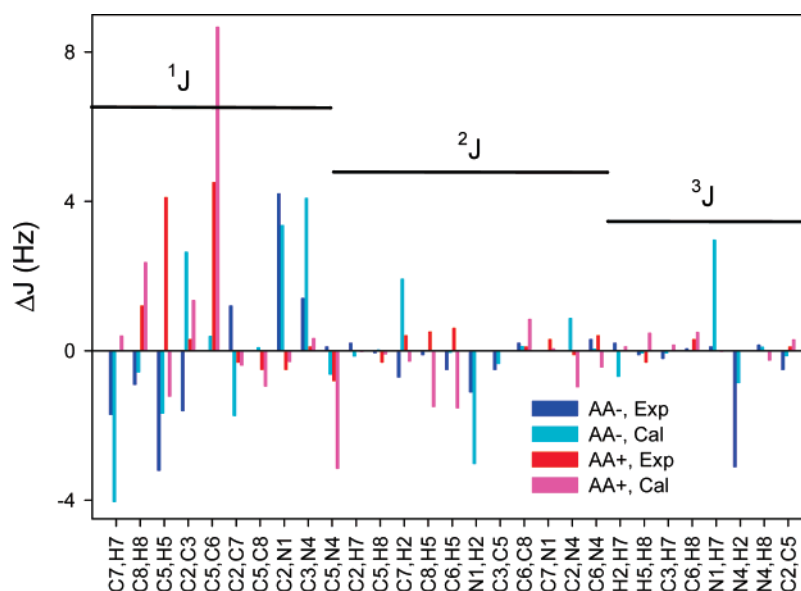


Figure 7. Computed and experimental pH-induced changes of spin–spin coupling constants (for the lowest-energy conformer of AA^+ and AA^- , with respect to that of the zwitterionic form AA^{ZW}).

TABLE 4: Spin–Spin Coupling Constants (Hz) Calculated for 10 AA[−] Conformers (A–F)^a and Comparison with Experimental Values

η^b	A	A'	A''	B	C	C'	C''	D	E	F	avg ^c	exptl ^d
J^1:												
C7,H7	125.0	124.7	125.0	125.0	125.5	124.9	125.5	124.9	123.9	123.7	125.0	129.0
C8,H8	125.8	125.8	125.7	125.0	125.7	125.6	125.8	124.8	125.9	125.1	125.7	128.3
C2,H2	135.9	136.8	134.0	135.4	137.7	130.6	137.8	130.5	136.2	135.1	135.7	142.2
C5,H5	139.6	139.6	139.5	127.6	139.2	138.8	139.6	127.6	140.1	127.6	139.4	139.5
C2,C3	48.3	53.4	53.4	47.8	52.5	52.6	47.1	52.5	48.8	48.2	50.5	50.5
C5,C6	52.3	52.3	52.4	52.3	52.5	52.4	52.1	52.5	51.3	52.4	52.3	54.4
C2,C7	36.5	31.3	35.9	36.0	34.0	35.3	35.2	35.6	36.6	36.6	35.0	35.2
C5,C8	34.6	34.7	34.6	39.2	34.6	34.9	35.0	39.0	35.2	39.2	34.7	34.9
C2,N1	−1.2	−1.0	−2.6	−1.2	−2.6	−2.0	−1.2	−2.2	−0.7	−0.8	−1.6	−4.5
C3,N4	−13.6	−14.6	−14.5	−13.8	−16.0	−15.9	−15.4	−16.7	−13.7	−13.9	−14.7	−15.7
C5,N4	−9.2	−9.3	−9.3	−9.0	−9.9	−9.9	−9.6	−9.8	−9.3	−9.1	−9.5	−10.6
J^2:												
C2,H7	−2.8	−2.8	−2.8	−2.8	−2.9	−3.1	−2.9	−3.1	−2.9	−3.0	−2.9	−4.2
C5,H8	−2.8	−2.8	−2.8	−3.3	−2.8	−2.8	−2.8	−3.3	−2.9	−3.3	−2.8	−4.25
C7,H2	−3.3	−0.4	−0.8	−3.5	−3.1	−3.5	−4.8	−3.8	−2.9	−3.1	−3.0	−5.0
C8,H5	−3.2	−3.2	−3.2	−3.8	−3.2	−3.2	−3.2	−3.5	−3.0	−3.8	−3.2	−4.8
C3,H2	−1.1	−3.9	−0.8	−1.1	−3.2	−6.1	−3.6	−5.9	−4.5	−4.2	−2.7	−5.0
C6,H5	−4.7	−4.7	−4.7	−5.9	−4.6	−4.6	−4.7	−5.9	−4.6	−5.9	−4.7	−5.3
N1,H2	−3.3	−3.9	0.7	−3.3	1.2	−1.2	−4.4	−0.7	−2.4	−2.4	−2.0	−1.1
C3,C5	−0.9	−0.6	−0.5	−0.8	−0.3	−0.3	−0.5	−0.1	−1.0	−0.9	−0.6	−0.5
C6,C8	−0.6	−0.6	−0.5	0.7	−0.6	−0.5	−0.5	0.7	−0.8	0.6	−0.6	1.2
C2,N4	−9.1	−9.3	−9.6	−9.3	−6.6	−6.4	−7.5	−6.5	−7.8	−7.6	−8.1	−6.8
C6,N4	−0.5	−0.5	−0.5	0.2	−0.7	−0.6	−0.6	0.2	−0.6	0.2	−0.6	−1.2
J^3:												
H2,H7	6.7	6.2	6.2	6.8	6.8	6.8	7.1	6.9	7.0	6.9	6.7	7.3
H5,H8	6.8	6.8	6.8	7.0	6.8	6.9	6.8	7.0	6.8	7.0	6.8	7.2
C3,H7	4.1	4.1	4.5	4.0	4.3	3.9	4.0	4.0	4.0	4.0	4.1	4.2
C6,H8	3.9	3.9	3.9	3.5	3.9	3.9	3.9	3.5	4.0	3.5	3.9	4.25
N1,H7	−3.7	−0.8	−3.4	−3.7	−2.9	−1.0	−3.7	−1.1	−3.8	−3.7	−2.8	−3.0
N4,H2	−0.8	−1.2	−1.2	−0.7	−0.7	−0.6	−1.1	−0.4	−1.2	−1.1	−0.8	−2.1
N4,H8	−2.5	−2.5	−2.5	−2.5	−2.5	−2.5	−2.5	−2.4	−2.5	−2.5	−2.5	−2.95
C2,C5	1.7	2.0	2.0	2.1	1.3	1.3	1.4	1.7	1.3	1.6	1.6	1.5
ΔJ^e	1.6	1.7	1.7	2.1	1.2	1.4	1.2	1.9	1.5	2.0	1.3	0.0

^a See Table 1 for the definition of the conformers. ^b Conformer ratios were estimated from the Boltzmann factor at 300 K. ^c The Boltzmann-weighted average. ^d pH = 12. ^e Average absolute deviations.

(−91.5, −91.4, and −91.6 Hz for the three rotamers), ² J (C5,H4) (3.2, 3.2, and 3.1 Hz), or ³ J (H4,H5) (7.1, 7.4, and 7.2 Hz) are rather insensitive to the rotation.

Also the computed coupling constants of the AA⁺ form listed in Table 5 agree best with the experiment for the energy-preferred conformer **A**. Average deviations of the **C** conformer are low as well, but this form can be excluded on the basis of the energy estimation. For the cation, especially the vicinal couplings (³ J) seem to be computed with significant errors. Fortunately, for AA⁺, the number of conformers that can contribute to the averaged observed J 's is smaller than for AA[−]. In fact, the conformer **A** seems to be clearly dominant (~90%), with Boltzmann population of the conformer **B** estimated from the equilibrium geometry being about 10%. Given the narrow potential well (Figure 2) in comparison with **A**, the population of conformer **B**, obtained through a complete integration over the two-dimensional potential energy surface, would be even smaller. Additionally, the average error ($\Delta J = 1.9$ Hz, see Table 5) of the couplings for the **B** form is much higher than that for **A** ($\Delta J = 1.4$ Hz).

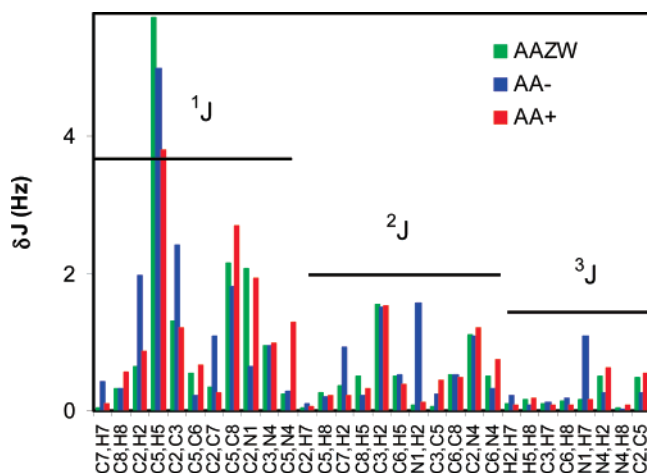
Finally, we can focus our attention on the conformational sensitivity of individual spin–spin coupling constants represented by the average absolute deviations plotted in Figure 8. In spite of the errors of the computed couplings, some constants clearly exhibit larger variations under the conformational change, and thus the NMR technique along with the quantum computation may be able to discriminate between peptide conformers. The ¹ J (C5,H5) coupling is the most sensitive one, particularly

to the change of the close φ angle, and can thus monitor the ratios of the (**A**, **C**) and (**B**, **D**) conformer classes (cf. Tables 4 and 5). The ψ torsion has little impact on this coupling. Similarly, the analogous ¹ J (C2,H2) coupling constant is more sensitive to the ψ angle, which makes it possible for these two couplings alone in principle to determine the AA secondary structure. To be able to generalize the results, however, one has to realize also the dependence on the molecular charge (the experimental ¹ J (C5,H5) constant is 146.8, 142.7, and 139.5 Hz for the cation, zwitterion, and anion, respectively), and in larger peptides similar variation can be expected for various amino acid side chains. The ¹ J (C5,C6) coupling, for example, is predominantly driven by the molecular charge (the measured values for the cation, zwitterion, and anion are 58.9, 54.4, and 54.4 Hz, which are actually very well reproduced by the calculation as 60.6, 51.9, and 52.3 Hz; see Tables 4 and 5 and ref 11), and the conformational variance does not exceed 0.7 Hz.

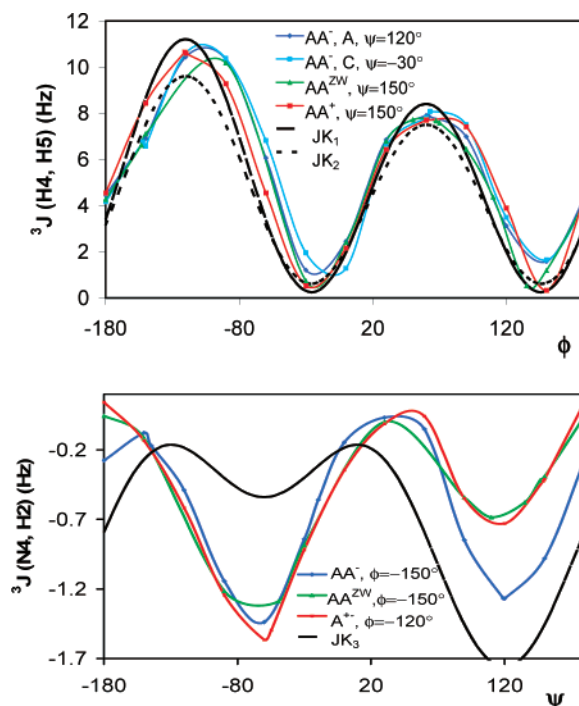
The three-bond couplings, however, are more important for the peptide structural determination than the one- and two-bond interactions.^{1,2,12,13} Therefore, their dependence on other factors than the torsion angle, such as charges of close molecular groups, is of paramount importance for peptide chemistry. As an example, two constants, ³ J (H4,H5) and ³ J (N4,H2), were selected and their dependence on the main-chain torsion angle plotted in Figure 9. Calculated curves for the AA cation, anion, and zwitterions are compared to the empirical Karplus-type curves derived in the literature on the basis of theoretical and

TABLE 5: Spin–Spin Coupling Constants (Hz) Calculated for Four AA⁺ Conformers and a Comparison with the Experiment^a

η	A	B	C	D	exptl
¹ J:	0.91	0.09	0.00	0.00	
C7,H7	129.1	129.2	128.9	128.9	130.7
C8,H8	128.7	127.7	128.7	127.5	130.4
C5,H5	140.1	132.5	140.8	133.2	146.8
C2,C3	52.1	51.1	54.6	53.4	52.4
C5,C6	60.6	59.2	60.8	59.5	58.9
C2,C7	32.7	32.7	32.1	32.2	33.7
C5,C8	33.7	38.9	33.5	39.1	34.4
C2,N1	−4.6	−4.6	−0.7	−0.8	−9.2
C3,N4	−18.4	−17.6	−16.4	−15.6	−17.0
C5,N4	−11.8	−9.4	−11.8	−9.1	−11.5
² J:					
C2,H7	−2.6	−2.6	−2.5	−2.5	−4.4
C5,H8	−2.9	−3.3	−2.9	−3.3	−4.5
C7,H2	−2.6	−2.7	−3.1	−3.1	−3.9
C8,H5	−4.7	−4.2	−4.7	−3.9	−4.2
C6,H5	−6.2	−7.2	−6.3	−6.8	−4.2
C6,C8	0.2	1.1	0.1	1.1	1.1
C7,N1	−0.5	−0.5	−0.4	−0.4	1.3
C2,N4	−11.1	−10.9	−8.7	−8.5	−6.9
C6,N4	−1.0	0.3	−1.2	0.5	−1.1
³ J:					
H2,H7	7.0	7.0	7.2	7.2	7.1
H5,H8	7.3	7.0	7.3	6.9	7.0
C3,H7	4.3	4.2	4.5	4.4	4.4
C6,H8	4.4	4.5	4.4	4.5	4.5
N1,H7	−3.8	−3.8	−4.1	−4.1	−3.1
N4,H8	−2.8	−2.6	−2.7	−2.7	−3.1
C2,C5	2.4	2.6	1.3	1.5	2.1
ΔJ	1.4	1.9	1.5	2.0	

^a The symbols have the same reference as in Table 4.**Figure 8.** Computed sensitivities of selected spin–spin coupling constants to the conformational change (average absolute deviations from the average over individual conformers A–F are plotted for each form).

experimental data. Clearly, for $^3J(\text{H4},\text{H5})$, the angular dependence does not rely significantly on the molecular charge and nicely corresponds to the two approximations proposed previously. Within $\varphi \approx -180 \dots -30^\circ$, the cationic AA⁺ $^3J(\text{H4},\text{H5})$ curve somewhat deviates from the anion and zwitterion, which can be explained by the vicinity of the COO[−] group, protonated in the cation. Some dispersion occurs also around $\varphi \approx 150^\circ$; this is, however, rather minor with respect to the principal conformational dependence.

**Figure 9.** Calculated dependence of the $^3J(\text{H4},\text{H5})$ (top) and $^3J(\text{N4},\text{H2})$ (bottom) vicinal spin–spin coupling constants on the encompassed torsion angle for the three charged AA forms and a comparison to the JK₁,³⁵ JK₂,³⁶ and JK₃³⁷ semiempirical Karplus curves.

Apparently, the $^3J(\text{N4},\text{H2})$ coupling (lower part of Figure 9) is more sensitive to the molecular charge and deviates more from the previously proposed curve. As expected, the anion AA[−] curve deviates from the cation and zwitterion because of the deprotonation of the NH₃⁺ group close to the rotating bond. However, a closer inspection reveals that the absolute coupling dispersion is like for the previous case, since the $^3J(\text{N4},\text{H2})$ constant varies in a much narrower range than $^3J(\text{H4},\text{H5})$.

Conclusions

On the basis of the DFT computations (BPW91/6-311++G**) of the two-dimensional potential energy surfaces, we were able to estimate the conformational behavior of the AA dipeptide under the pH changes. Whereas the neutral zwitterionic form AA^{ZW} and the cation (AA⁺) adopt similar conformations of the main peptide chain (φ , ψ angles), the anion (AA[−]) exists in two forms differing by the ψ -angle values. The forms are approximately equally populated in aqueous solutions at room temperature. The anion main chain folding is more complex than for the other forms because of the influence of the NH₂ group, which can serve both as a hydrogen donor and acceptor in an intramolecular hydrogen bonding. The results of the analysis of the potential energy surfaces are in agreement with both experimental and calculated NMR chemical shifts and spin–spin coupling constants. The NMR parameters could be calculated with a limited accuracy, but the pH dependence of the chemical shifts for the dipeptide as well as for the alanine monomer could be explained on the basis of the theory. Furthermore, the comparison of the experimental and calculated coupling constants is consistent with the energetic analysis.

Acknowledgment. The work was supported by the Czech Science Foundation (Grant Nos. 203/06/0420 and 202/07/0732), the Grant Agency of the Academy of Sciences (A4005507020, A400550701), and the Ministry of Education, Youth and Sports (Grant No. LC512).

Supporting Information Available: Cartesian coordinates, absolute energies of equilibrium structures, and additional computational and experimental details. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Kessler, H.; Seip, S. In *Two-Dimensional NMR Spectroscopy, Application for Chemists and Biochemists*; Croasmun, W. R., Carlson, R. M. K., Eds.; VCH Publ.: New York, 1994; p 619.
- (2) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
- (3) Englander, S. W.; Mayne, L. *Annu. Rev. Biophys. Biomol. Struct.* **1992**, *21*, 243.
- (4) Evans, J. N. S. *Biomolecular NMR Spectroscopy*; Oxford University Press: Oxford, 1995.
- (5) William, M. P.; Waltho, J. P. *Chem. Soc. Rev.* **1992**, *21*, 227.
- (6) Wolinski, K.; Hilton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.
- (7) Sychrovský, V.; Gräfenstein, J.; Cremer, D. *J. Chem. Phys.* **2000**, *113*, 3530.
- (8) Helgaker, T.; Watson, M.; Handy, N. C. *J. Chem. Phys.* **2000**, *113*, 9402.
- (9) Barone, V.; Peralta, J. E.; Contreras, R. H.; Snyder, J. P. *J. Phys. Chem. A* **2002**, *106*, 5607.
- (10) Bouř, P.; Sychrovský, V.; Maloň, P.; Hanzlíková, J.; Baumruk, V.; Pospíšek, J.; Buděšínský, M. *J. Phys. Chem. A* **2002**, *106*, 7321.
- (11) Bouř, P.; Buděšínský, M.; Špirko, V.; Kapitán, J.; Šebestík, J.; Sychrovský, V. *J. Am. Chem. Soc.* **2005**, *127*, 17079.
- (12) Karplus, M.; Weaver, D. L. *Nature* **1976**, *260*, 404.
- (13) Perere, S. A.; Bartlett, R. J. *Magn. Reson. Chem.* **2001**, *39*, S183.
- (14) Laws, D. D.; deDios, A. C.; Oldfield, E. J. *Biomol. NMR* **1993**, *3*, 607.
- (15) Birn, J.; Poon, A.; Mao, Y.; Ramamoorthy, A. *J. Am. Chem. Soc.* **2004**, *126*, 8529.
- (16) Sebastiani, D.; Rothlisberger, U. *J. Phys. Chem. B* **2004**, *108*, 2807.
- (17) Čejchan, A.; Špirko, V. *J. Mol. Spectrosc.* **2003**, *217*, 142.
- (18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (19) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (20) Barone, V.; Cossi, M.; Tomasi, J. *J. Comput. Chem.* **1998**, *19*, 404.
- (21) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
- (22) Ruud, K.; Helgaker, T.; Bak, K. L.; Jorgensen, P.; Jensen, H. J. A. *J. Chem. Phys.* **1993**, *99*, 3847.
- (23) Bouř, P.; Buděšínský, M. *J. Chem. Phys.* **1999**, *110*, 2836.
- (24) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (25) Kutzelnigg, W.; Fleischer, U.; Schindler, M. *NMR - Basic Principles and Progress*; Springer: Heidelberg, 1990; Vol. 23.
- (26) Podolsky, B. *Phys. Rev.* **1928**, *32*, 812.
- (27) Ixaru, L. *Numerical Methods for Differential Equations and Applications*; Reidel: Dordrecht, 1984.
- (28) Špirko, V.; Piecuch, P.; Bludský, O. *J. Chem. Phys.* **2000**, *112*, 189.
- (29) Creighton, T. E. *Proteins*; New York, 1984.
- (30) Fadda, E.; Casida, M. E.; Salahub, D. R. *J. Phys. Chem. A* **2003**, *107*, 9924.
- (31) Bouř, P. *J. Chem. Phys.* **2004**, *121*, 7545.
- (32) Kümmel, S.; Kronik, L.; Perdew, J. P. *Phys. Rev. Lett.* **2004**, *93*, 213002.
- (33) Bouř, P.; Raich, I.; Kaminský, J.; Hrabal, R.; Čejka, J.; Sychrovský, V. *J. Phys. Chem. A* **2004**, *108*, 6365.
- (34) *The Encyclopedia of Computational Chemistry*; Schleyer, P. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III, Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998.
- (35) Case, D. A.; Scheurer, C.; Brüschweiler, R. *J. Am. Chem. Soc.* **2000**, *122*, 10390.
- (36) Schmidt, J. M.; Blümel, M.; Löhr, F.; Rüterjans, H. *J. Biomol. NMR* **1999**, *14*, 1.
- (37) Demarco, A.; Llinas, M.; Wüthrich, K. M. *Biopolymers* **1978**, *17*, 2727.

Effect of local sugar and base geometry on ^{13}C and ^{15}N magnetic shielding anisotropy in DNA nucleosides

Eva Brumovská · Vladimír Sychrovský · Zuzana Vokáčová · Jiří Šponer · Bohdan Schneider · Lukáš Trantírek

Received: 12 September 2008 / Accepted: 17 September 2008 / Published online: 14 October 2008
© Springer Science+Business Media B.V. 2008

Abstract Density functional theory was employed to study the dependence of ^{13}C and ^{15}N magnetic shielding tensors on the glycosidic torsion angle (χ) and conformation of the sugar ring in 2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxycytidine, and 2'-deoxythymidine. In general, the magnetic shielding of the glycosidic nitrogens and the sugar carbons was found to depend on both the conformation of the sugar ring and χ . Our calculations indicate that the magnetic shielding anisotropy of the C6 atom in pyrimidine and the C8 atom in purine bases depends strongly on χ . The remaining base carbons were found to be insensitive to both sugar pucker and χ re-orientation. These results call into question the underlying assumptions of currently established methods for interpreting residual chemical shift anisotropies and ^{13}C and ^{15}N auto- and cross-correlated

relaxation rates and highlight possible limitations of DNA applications of these methods.

Keywords Magnetic shielding · Chemical shift anisotropy · DNA · RNA

Introduction

Knowledge of the magnitude and orientation of ^{13}C and ^{15}N magnetic shielding (MS)/chemical shift (CS) tensors in DNA and RNA nucleosides is essential for the interpretation of NMR relaxation data and for the analysis of residual chemical shift anisotropy (RCSA) resulting from weak alignment (Akke et al. 1997; Boisbouvier et al. 2000; Duchardt et al. 2004; Duchardt and Schwalbe 2005; Ferner et al. 2008; Grishaev et al. 2006; Hansen and Al-Hashimi 2006; Ravindranathan et al. 2003; Ravindranathan et al. 2005; Schofberger et al. 2006; Shajani and Varani 2007; Sychrovský et al. 2005; Trantírek et al. 2007). Large chemical shift anisotropies (CSAs) of ^{13}C and ^{15}N nuclei in nucleic acids have been found useful for constraining bases relative to the molecular alignment tensor (Grishaev et al. 2006; Hansen and Al-Hashimi 2006), glycosidic torsion angle (Duchardt et al. 2004), sugar ring conformation (Boisbouvier et al. 2000), and the evaluation of conformational dynamics around a glycosidic bond (Ravindranathan et al. 2003). Until quite recently, the only available experimental data on the ^{13}C and ^{15}N CS-tensor magnitude and orientation in nucleic acids originated from solid-state NMR measurements and were limited to model compounds such as nucleic acid bases (Hu et al. 1998) or nucleosides (Stueber and Grant 2002).

In a recent report, Duchardt and Schwalbe simultaneously analyzed relaxation measurements for all protonated base carbons in the nucleosides of a small RNA hairpin, using

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9278-7) contains supplementary material, which is available to authorized users.

E. Brumovská · L. Trantírek (✉)
Faculty of Science, University of South Bohemia and Biology
Centre AS CR v.v.i., Branišovská 31, 370 05 České Budějovice,
Czech Republic
e-mail: trant@paru.cas.cz

V. Sychrovský · Z. Vokáčová
Institute of Organic Chemistry and Biochemistry, AS CR v.v.i.,
Flemingovo náměstí 2, 166 10 Prague, Czech Republic

J. Šponer
Institute of Biophysics, AS CR v.v.i., Královopolská 135, 612 65
Brno, Czech Republic

B. Schneider
Biotechnological Institute AS CR, Vídeňská 1083, 142 20
Prague, Czech Republic

solid-state CSA values from Stueber and Grant (2002) (Duchardt and Schwalbe 2005). They found that the use of CSA values obtained from solid-state NMR measurements led to the remarkable result of systematically lower order parameters for the purine C8 atom compared to pyrimidine C6 sites, even for base paired nucleotides. They also reported that relaxation data recorded for ^{13}C and ^{15}N in the same nucleotide revealed a mismatch in the commonly used CSA values, and raised a question as to whether CSA values in solution NMR require adjustment from solid-state values.

Very recently, ^{13}C chemical shift anisotropies in right-handed double-helical DNA and RNA fragments were characterized in solution using the dependence of the relaxation rates on the magnetic field (Ying et al. 2006a), cross-correlated relaxation rates (Ravindranathan et al. 2005), and liquid crystal measurements (Bryce et al. 2005; Hansen and Al-Hashimi 2006; Ying et al. 2006b). Differences of up to 30 ppm were found between CSA values obtained from solid-state and solution NMR measurements. It was suggested that the different ^{13}C chemical shift anisotropies stemmed from differences in hydration of Watson–Crick base-paired oligonucleotides, molecular geometry, and electrostatic crystal potential (Stueber and Grant 2002; Ying et al. 2006b). It is generally recognized that the CS tensors of ^1H , ^{13}C , ^{15}N , and ^{31}P nuclei in nucleotides depend on local conformation (Dejaegere and Case 1998; Ebrahimi et al. 2001; Precechtelova et al. 2007; Sitkoff and Case 1998; Sychrovsky et al. 2005) and hydrogen bonding (Czernek et al. 2000). Dejaegere and Case (1998) reported calculations of MS tensors for methyl β -D-2'-deoxyribofuranoside and methyl β -D-ribofuranoside as models for the 2'-deoxyribose and ribose sugars in nucleic acids and found that the magnetic shielding anisotropies (MSA) for C1' and C3' are sensitive to puckering of the sugar ring. The dependence of the C1' and C3' MS tensors on sugar pucker in RNA polynucleotides was later confirmed experimentally by Boisbouvier et al. (2000). Recently, Sychrovsky et al. (2005) reported calculations of the C1' and N1/9 MS-tensors in DNA nucleosides. They demonstrated that both the magnitude and orientation of C1' and N1/9 MS tensors depend on the glycosidic torsion and sugar ring conformation and that accounting for conformation-dependent variability in these tensors may be crucial for proper interpretation of cross-correlated relaxation rates between the N1/9 CS-tensor and C1'–H1' dipole–dipole interaction in nucleic acids. Importantly, these studies indicated that variations in MSA values due to differences in local molecular geometry may be two to three times larger than those due to the environmental differences experienced by nucleic acids in the solid and solution states.

Recent improvements in spectrometer hardware and measurement strategies permit the MS-tensor-related NMR

parameters to be measured at levels of accuracy where their quantitative interpretation is limited by an unknown degree of site-to-site MS tensor variation. Due to a lack of information about ^{13}C and ^{15}N MS-tensors outside the right-handed double helical region, the quantitative interpretation of MS-tensor related NMR parameters in non-canonical regions, left-handed double helices, or multi-strand forms of DNA should be avoided as it might lead to structural errors (Sychrovsky et al. 2005). Mapping conformational dependencies of ^{13}C and ^{15}N MS-tensors in DNA is therefore valuable as it allows MS-tensor related NMR parameters to be used for structural analysis of biologically important nucleic acid motifs such as Z-DNA, DNA triplexes and quadruplexes, or unusual DNA duplex and hairpin motifs.

In the present study, we employed density functional theory (DFT) to investigate the relationships between the anisotropy of ^{13}C and ^{15}N MS and the sugar ring conformations and glycosidic torsion angle in 2'-deoxyadenosine, 2'-deoxyguanosine, 2'-deoxythymidine, and 2'-deoxycytidine. The main objectives were (1) to determine the range of individual principal components of the tensor for the experimentally observed sugar puckers and glycosidic bond orientation, and (2) to determine the changes in MS tensor orientations resulting from changes in sugar pucker and glycosidic bond conformation.

Methods

The compounds 2'-deoxyadenosine (dAde), 2'-deoxyguanine (dGua), 2'-deoxycytidine (dCyt), 2'-deoxythymidine (dThy) (Fig. 1) were used as models for all calculations of the ^{13}C and ^{15}N MS tensors. The geometry of all nucleosides was gradient optimized with the B3LYP exchange-correlation functional (Becke 1993; Lee et al. 1988) and the 6-31G(d,p) atomic basis set. In the initial geometry optimization, the χ torsion angle was estimated to be close to either *syn* or *anti* minima (the *anti* region was defined as: $180^\circ < \chi < 280^\circ$, and *syn* as $50^\circ < \chi < 80^\circ$), and the sugar was adjusted to either C3'-endo (Pseudorotation angle P (Altona and Sundaralingam 1972) set to approximately 20°) or C2'-endo (P about 160°). A constrained geometry optimization for a stepwise change in torsion angle χ was then performed for each nucleoside. All parameters were freely optimized except χ . The magnetic shielding tensors were calculated using the GIAO approach (Wolinski et al. 1990) with the B3LYP functional and the atomic basis set (9s,5p,1d/5s,1p)[6s,4p,1d/3s,1p] for both carbon and nitrogen (Kutzelnigg et al. 1991). All calculations were performed with the Gaussian G03 program (Frisch et al. 2004).

A second-rank MS tensor in the principal axis system was obtained from the NMR calculations. The full

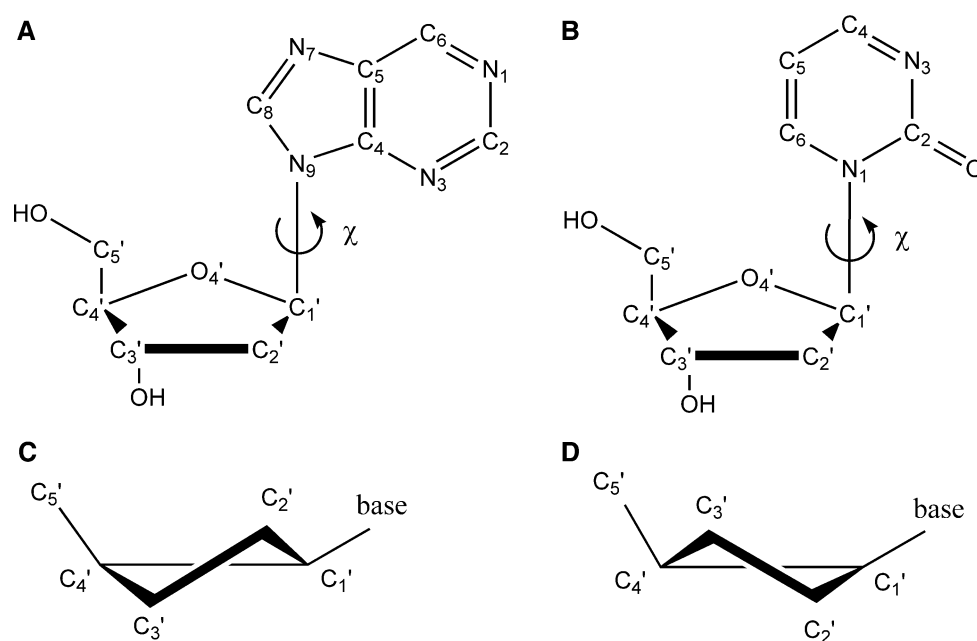


Fig. 1 Structure and atom numbering scheme for **a** 2'-deoxypurine and **b** 2'-deoxypyrimidine. Carbon and nitrogen atoms are numbered according to IUPAC nomenclature (Markley et al. 1998). The glycosidic torsion angle χ is defined by atoms O4'-C1'-N9-C4 in purines and O4'-C1'-N1-C2 in pyrimidines. There are two energetically favored regions of χ , *anti* and *syn*. In agreement with

comparative studies of crystallographic data, the *anti* region is defined as $180^\circ < \chi < 280^\circ$, and the *syn* region as $50^\circ < \chi < 80^\circ$. **c** and **d** Schematic representations of the two main sugar conformations in DNA, C2'-endo (Pseudorotation angle P (Altona and Sundaralingam 1972) approximately 160°) and C3'-endo (P about 20°)

MS tensor was decomposed into its isotropic and anisotropic (herein, referred to as MSA tensor) parts. A traceless MSA tensor in the principal axis system is described by two adjustable parameters: its magnitude $|\sigma^M| = (\sigma_x^2 + \sigma_y^2 - \sigma_x \sigma_y)^{1/2}$, where $\sigma_x = \sigma_{11} - \sigma_{33}$ and $\sigma_y = \sigma_{22} - \sigma_{33}$, and its asymmetry, $\eta = (\sigma_{22} - \sigma_{11})/\sigma_{33}$, where $\sigma_{11} \leq \sigma_{22} \leq \sigma_{33}$.

Results

Dependence of MSA on sugar pucker and glycosidic torsion angle

Base nitrogens

The calculated nitrogen MSA tensor magnitudes and orientations as a function of the sugar pucker and glycosidic torsion angle (χ) for pyrimidine and purine 2'-deoxynucleosides are displayed in Figs. 2 and 3, respectively. The MSA tensors of the base nitrogen atoms are almost independent of the sugar pucker with the exception of the N1 pyrimidine and N9 purine atoms involved in the glycosidic bond. Their MSA tensors depend on both orientation of the glycosidic bond and sugar pucker. While the magnitudes $|\sigma^M|$ of the glycosidic nitrogen MSA tensors in

2'-deoxypyrimidines in the *syn* conformation are essentially the same for the C2'- and C3'-endo sugar puckers, our calculations indicate that sugar pucker mode has a pronounced influence on $|\sigma^M|$ values in the *anti* region. For *anti* 2'-deoxypyrimidines with C3'-endo sugars, the absolute values of $|\sigma^M|$ are expected to be 10–12 ppm larger than $|\sigma^M|$ values for *anti* 2'-deoxypyrimidines with C2'-endo sugars. For 2'-deoxythymidine, asymmetry of the glycosidic nitrogen MSA tensor is independent of the conformation of the sugar ring. In contrast, asymmetry of the tensor in 2'-deoxycytidine depends on sugar pucker mode. For both *syn* and *anti* 2'-deoxycytidine with C3'-endo sugars, the absolute values of η are predicted to be smaller by 0.2–0.25 than those of 2'-deoxycytidine with C2'-endo sugars. In general, the conformation of the glycosidic bond and sugar pucker has only a moderate effect on the orientation of the glycosidic nitrogen MSA tensor. The most pronounced effect has been found for the N1 atom of 2'-deoxycytidine in the *anti* conformation. The orientation of σ_{11} , in *anti* 2'-deoxycytidine with a C2'-endo sugar differs from the orientation observed in *anti* 2'-deoxycytidine with a C3'-endo sugar by $\sim 15^\circ$.¹

¹ The data for the N1/9 were obtained in our previous study (supplementary material) (Sychrovsky et al. 2005).

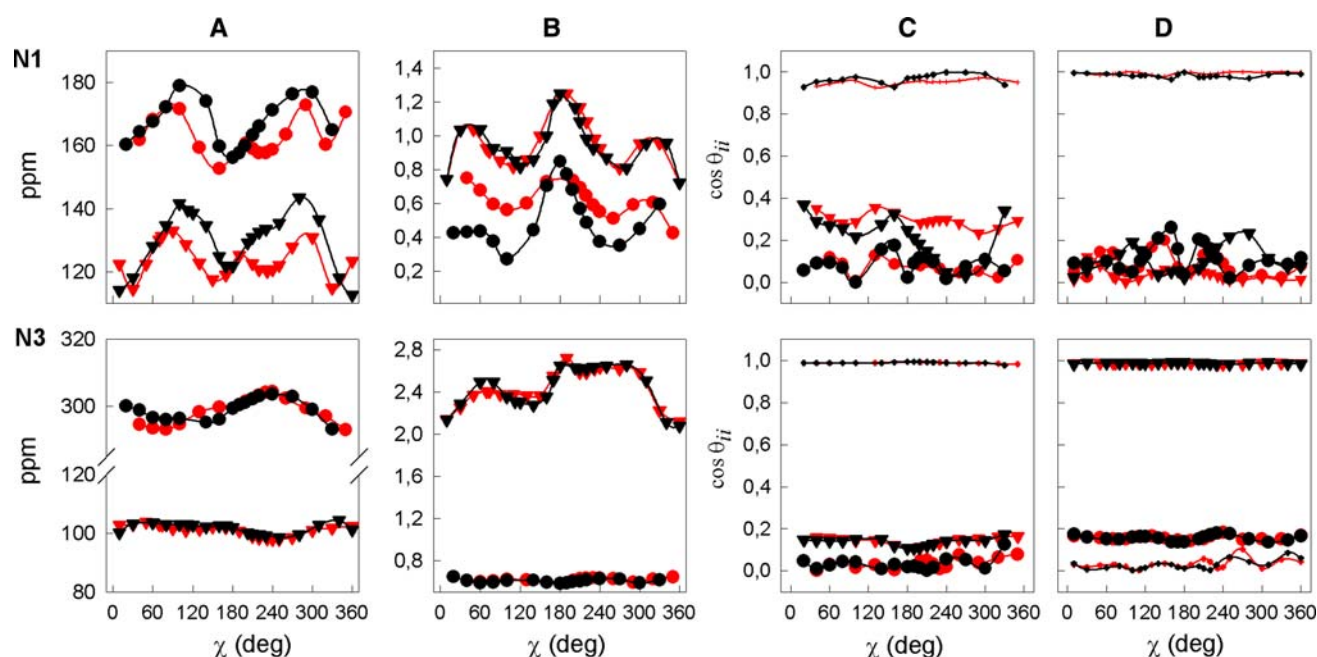


Fig. 2 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of ^{15}N MSA tensors for 2'-deoxycytidine (—●—) and 2'-deoxythymidine (—▼—) as a function of glycosidic torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation (θ_{11} (▼), θ_{22} (●), θ_{33} (●)) of individual components of the MSA tensor with respect to the N1–C2 bond as a function of torsion

angle χ in C2'-endo (in red) and C3'-endo (in black) sugar conformation in 2'-deoxycytidine and 2'-deoxythymidine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$. The data for N1 were obtained in our previous study (supplementary material) (Sychrovsky et al. 2005)

In 2'-deoxypurines, the sugar pucker mode has negligible influence on the $|\sigma^M|$ or η values and orientations of the N1, N3, and N7 MSA tensors (Fig. 3). However, analogous to the 2'-deoxypyrimidines, sugar pucker mode strongly influences the glycosidic nitrogen MSA tensor magnitude and shape as well as its orientation with respect to the base geometry. While the magnitudes of the tensor in *syn* 2'-deoxypurines are independent of sugar pucker mode, the differences in $|\sigma^M|$ values between C2'-endo and C3'-endo sugar pucker reaches up to 12 ppm in the *anti* region. For *syn* 2'-deoxypurines with C2'-endo sugar pucker, the absolute value of η is reduced by 0.1 compared to 2'-deoxypurines with C3'-endo sugar pucker. For *anti* 2'-deoxypurines, the corresponding difference ranges up to 0.35. In addition, our calculations indicate that when sugar pucker changes from C2'-endo to C3'-endo, the glycosidic nitrogen MSA tensor reorients itself by about 15°.

Base carbons

The calculated base carbon MSA tensor magnitudes, asymmetries and orientations as a function of the sugar pucker and glycosidic torsion angle for pyrimidine and purine 2'-deoxynucleosides are depicted in Figs. 4 and 5, respectively. In general, the MSA tensors of the base carbon atoms exhibited negligible dependence on sugar

pucker. As expected, our analysis indicated that the MSA tensors of the quaternary 2'-deoxypyrimidine C4 atoms are not influenced by the orientation of the glycosidic bond, in contrast to the carbon atoms proximal to the glycosidic linkage (i.e. C2, C6 and to some extent C5). While differences in $|\sigma^M|$ of the C5 atom MSA tensors between *syn* and *anti* conformations are only about 5 ppm, the corresponding difference for C6 atoms can be as great as 20–25 ppm. In 2'-deoxypyrimidines, the C2 MSA tensor $|\sigma^M|$ and η values are modulated by the orientation of the glycosidic bond (Fig. 4). The differences in absolute values of $|\sigma^M|$ and η between *syn* and *anti* conformations are about 10 ppm and 0.3. In addition, the C2 MSA tensor asymmetry η is influenced by the conformation of the sugar ring. However, the absolute η values for C2'-endo and C3'-endo sugar pucker are essentially the same in both *syn* and *anti* regions. Our calculations indicate that the η and $\cos \sigma_{ii}$ values of the C4 and C5 MSA tensors are quite insensitive to reorientation of the glycosidic torsion angle. In contrast to the C6 MSA tensor orientation, the asymmetry of the tensor is substantially modulated by χ . The difference between C6 η values corresponding to the *syn* and *anti* conformations is about 0.5.

The MSA tensors of the base carbons in 2'-deoxypurines appeared to be essentially independent of χ with the exception of the C8 atoms (Fig. 5). Although the

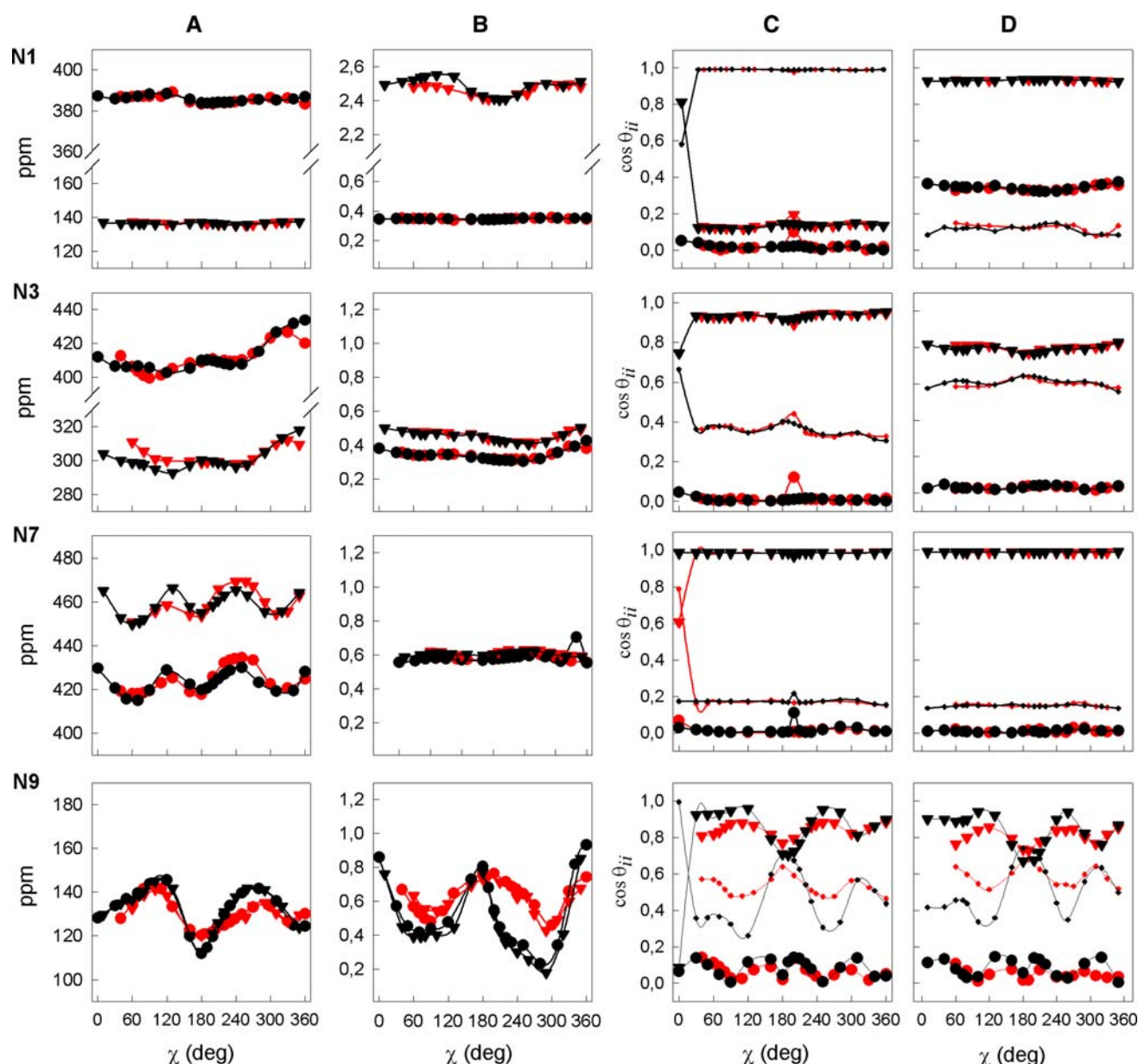


Fig. 3 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of ^{15}N MSA tensors for 2'-deoxyadenine (—●—) and 2'-deoxyguanosine (—▼—) as a function of glycosidic torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation (θ_{11} (▼), θ_{22} (●), θ_{33} (●)) of individual components of the MSA tensor with respect to the N9–C4 bond as a function of torsion

angle χ in C2'-endo (in red) and C3'-endo (in black) sugar conformation in 2'-deoxyadenosine and 2'-deoxyguanosine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$. The data for N9 were obtained in our previous study (supplementary material) (Sychrovsky et al. 2005)

conformation of χ clearly does not influence orientation of the C8 MSA tensor, it has a pronounced influence on both $|\sigma^M|$ and η . For 2'-deoxyadenosine, the differences in the absolute values of $|\sigma^M|$ and η between the *syn* and *anti* conformation are about 12 ppm and 0.6, respectively. For 2'-deoxyguanosine, the corresponding differences are about 10 ppm and 0.8. Although the C4 and C5 MSA tensor $|\sigma^M|$ and η values appear to be modulated by orientation of the glycosidic bond, the absolute values of

$|\sigma^M|$ and η are essentially the same in the stereochemically important regions of χ . The orientations of the C4 MSA tensors in 2'-deoxypurines are independent of χ . It is worth mentioning (see “Discussion”) that while the orientation of the C5 MSA tensor in 2'-deoxyguanosine is independent of the glycosidic bond angle, our calculations indicate that the C5 MSA tensor of 2'-deoxyadenosine differs slightly in orientation between the *syn* and *anti* regions.

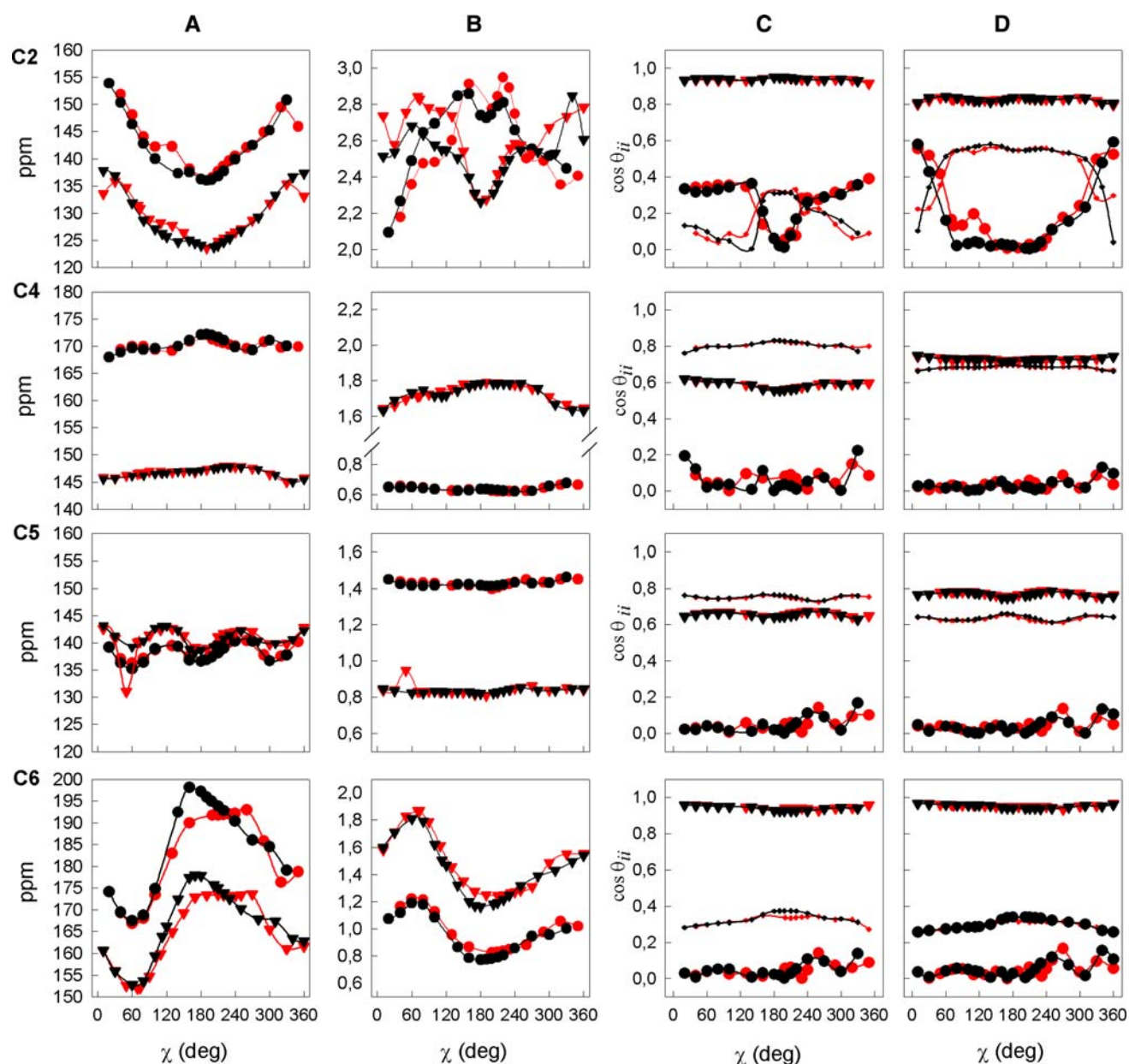


Fig. 4 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of base ^{13}C MSA tensors for 2'-deoxycytidine (—●—) and 2'-deoxythymidine (—▼—) as a function of glycosidic torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation (θ_{11} (▼), θ_{22} (●), θ_{33} (●)) of individual components of the

MSA tensor with respect to the N1–C2 bond as a function of torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation in 2'-deoxycytidine and 2'-deoxythymidine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$

Sugar carbons

Figures 6 and 7 depict the calculated magnitudes, asymmetries, and orientations for the MSA tensors of sugar carbons in 2'-deoxypyrimidines and 2'-deoxypurines, respectively. As anticipated, the MSA tensors of all endocyclic sugar carbons are significantly affected by the sugar pucker mode. In 2'-deoxypyrimidines, the orientation of the glycosidic torsion angle has only a small influence on the MSA tensor with exception of the C1' and C2' carbon

atoms. Noteworthy, the $|\sigma^M|$ of the C1' MSA changes up to 25 ppm upon re-orientation of the torsion angle χ between *syn* and *anti* regions.² The quantity most sensitive to the sugar pucker mode is the magnitude of the C3' carbon MSA tensor. The difference in the absolute value of $|\sigma^M|$ between C2'- and C3'-endo sugar conformations can be up

² The data for C1' were obtained in our previous study (supplementary material) (Sychrovsky et al. 2005).

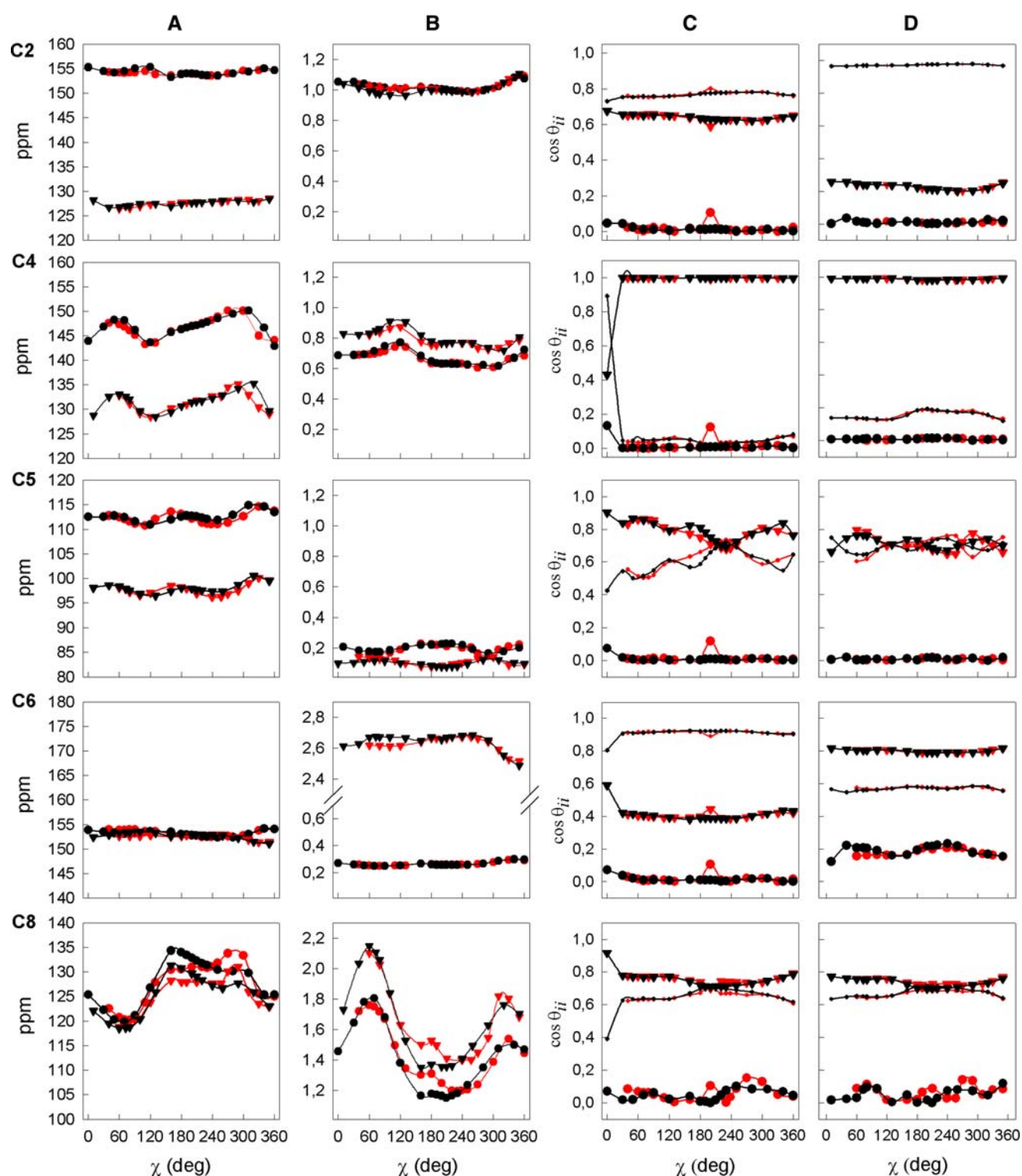


Fig. 5 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of base ^{13}C MSA tensors for 2'-deoxyadenosine (—●—) and 2'-deoxyguanosine (—▼—) as a function of glycosidic torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation (θ_{11} (▼), θ_{22} (●), θ_{33} (●)) of

individual components of the MSA tensor with respect to the N9–C4 bond as a function of torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation in 2'-deoxyadenosine and 2'-deoxyguanosine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$

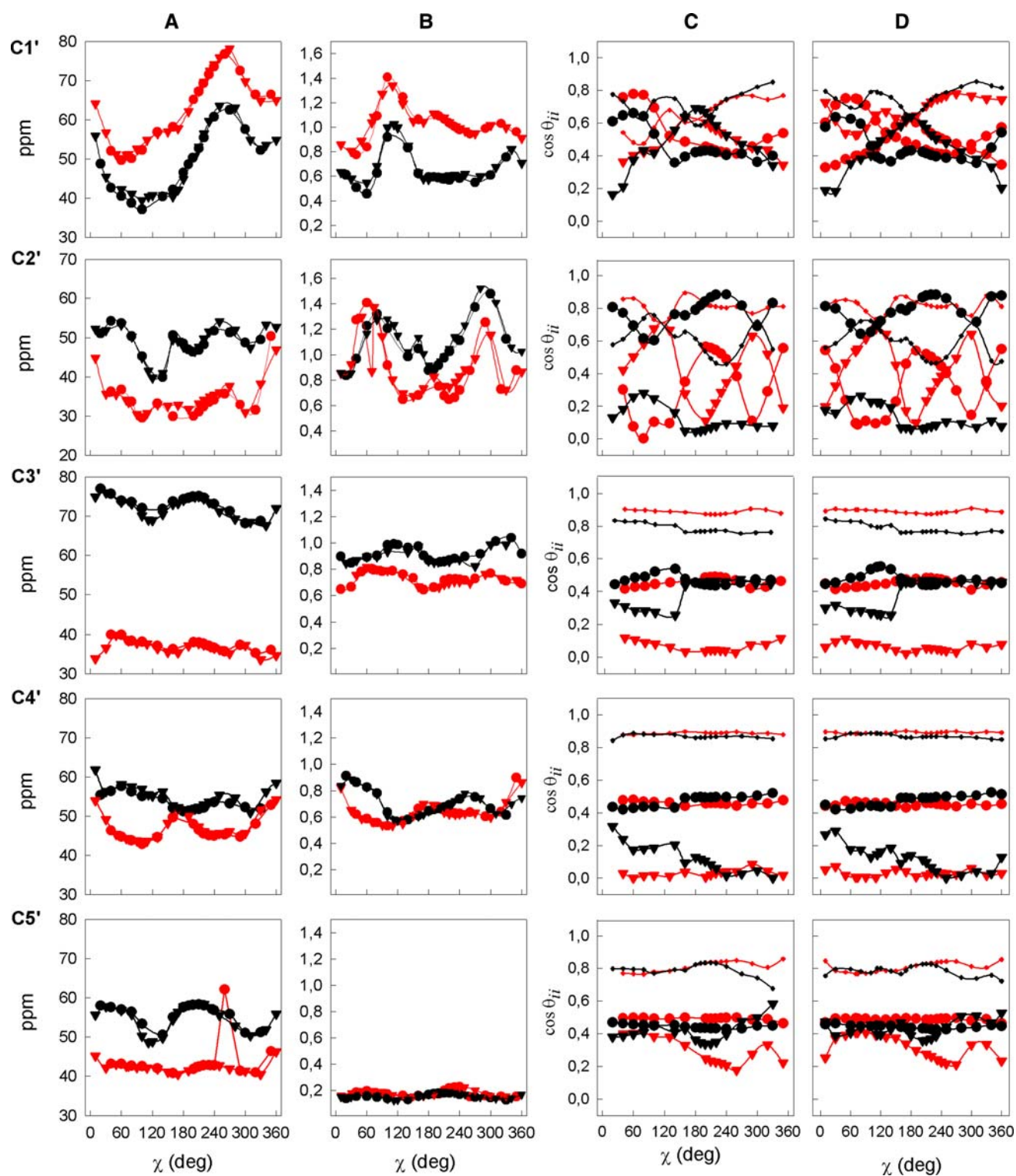


Fig. 6 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of sugar ^{13}C MSA tensors for 2'-deoxycytidine (\bullet) and 2'-deoxythymidine (\blacktriangledown) as a function of glycosidic torsion angle χ in C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation ($\theta_{11}(\blacktriangledown), \theta_{22}(\bullet), \theta_{33}(\bullet)$) of individual

components of the MSA tensor for C1', C2', C3', C4', and C5' carbons with respect to the C1'–N1/9, C1'–C2', C2'–C3', C3'–C4', and C4'–C5' bonds for 2'-deoxycytidine and 2'-deoxythymidine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$. See footnote 2

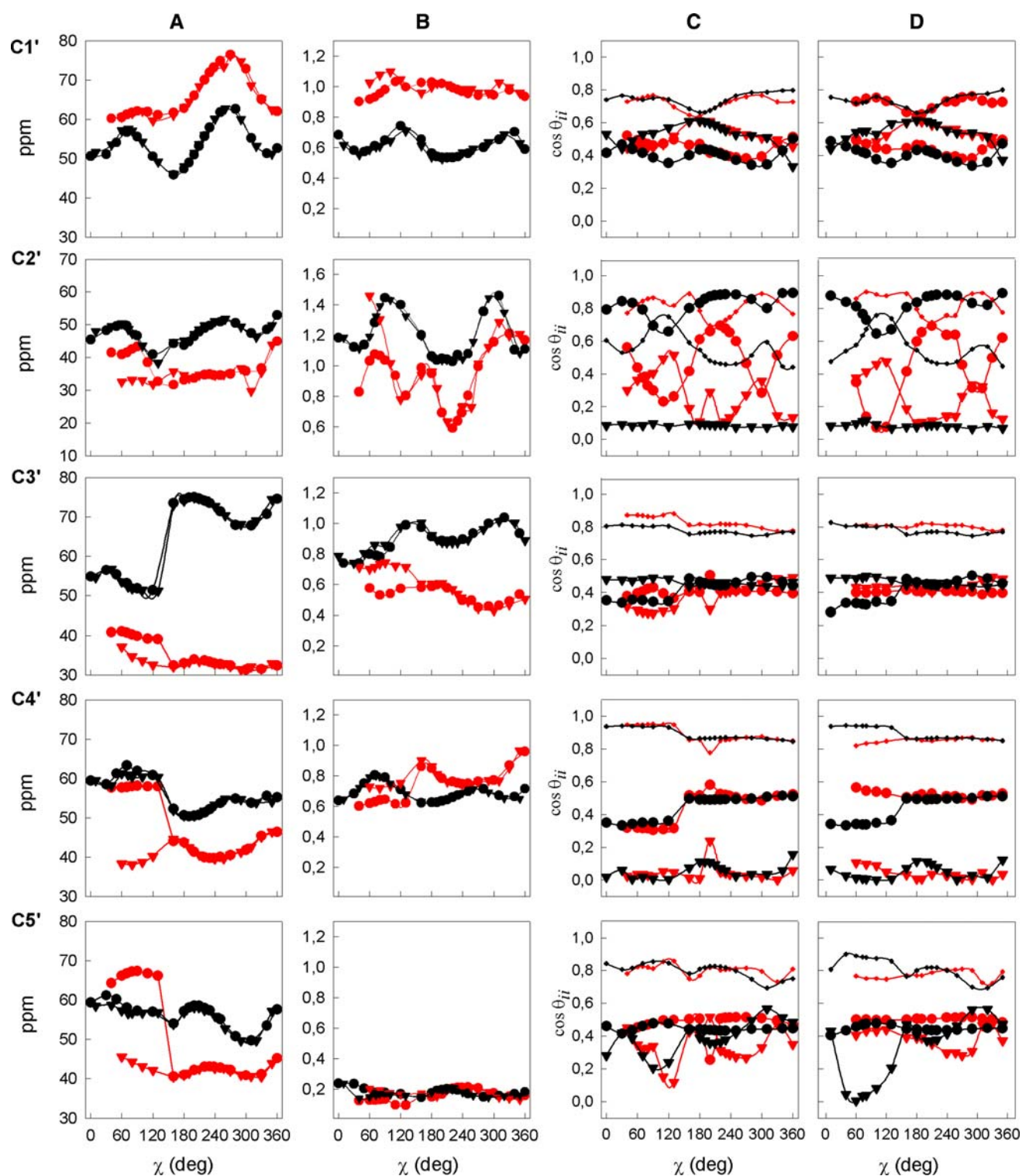


Fig. 7 Calculated magnitudes $|\sigma^M|$ (a) and asymmetries η (b) of sugar ^{13}C MSA tensors for 2'-deoxyadenosine (—●—) and 2'-deoxyguanosine (—▼—) as a function of glycosidic torsion angle χ in the C2'-endo (in red) and C3'-endo (in black) conformation. c and d display the calculated orientation (θ_{11} (▼), θ_{22} (●), θ_{33} (●)) of individual

components of the MSA tensor for C1', C2', C3', C4', and C5' carbon atoms with respect to the C1'-N1/9, C1'-C2', C2'-C3', C3'-C4', and C4'-C5' bonds for 2'-deoxyadenosine and 2'-deoxyguanosine, respectively. The orientations are expressed in absolute values of $\cos \theta_{ii}$. See footnote 2

to 35 ppm. The corresponding differences for the other sugar carbon atoms range from 10 to 22 ppm.

The MSA tensors of the endo-cyclic sugar carbons in 2'-deoxypurines display a pattern of conformational dependence similar to the 2'-deoxypyrimidines. However, in contrast to the 2'-deoxypyrimidines, the $|\sigma^M|$ values of the C3', C4', and C5' MSA tensors in 2'-deoxypurines are sensitive to reorientation of the glycosidic torsion angle. In *syn* 2'-deoxyguanosine with C3'-endo sugar, the N3 gets close to H3' (~ 2.5 Å). This might affect the C3' MSA. Similarly, in *syn* 2'-deoxypurines with C2'-endo sugars, the N3 get close to the H5' (~ 2.8 Å). In this case, a weak interaction between electron cloud from the N3 and H5' might affect the C5' MSA. The C5' MSA is sensitive to reorientation of the glycosidic torsion angle only in 2'-deoxyguanosine, but not in 2'-deoxyadenosine. The explanation for this may be in pronouncedly lower electron density at the N3 of 2'-deoxyadenosine as compared to the N3 of 2'-deoxyguanosine. The C4' MSA seems to be affected indirectly. The effect appears to be coupled with weak contact formation between the N3 of 2'-deoxyguanosine and its H5'.

Basis set dependence of DFT calculations

The choice of atomic basis for MS calculation may affect the accuracy of the modeled MSA tensor (Schoffberger et al. 2006; Sitkoff and Case 1998; Stueber and Grant 2002). To estimate the basis set effect, we performed calculations of the $|\sigma^M|$ with the Iglo II and a larger Iglo III basis for the selected grid point geometries of 2'-deoxyguanosine (Fig. 8). For *syn* ($\chi = 40^\circ \dots 110^\circ$, supplementary material) 2'-deoxyguanosine with C2'-endo sugar pucker, the absolute

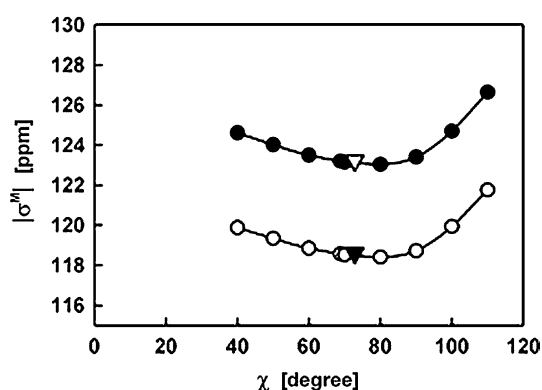


Fig. 8 Comparison of the $|\sigma^M|$ values for C8 MSA tensors calculated using the Iglo II (open circles) and Iglo III (filled circles) basis sets. The values were calculated for 2'-deoxyguanosine with C2'-endo sugar pucker and $\chi = 40^\circ, 50^\circ, 60^\circ, 68.8^\circ, 70^\circ, 80^\circ, 90^\circ, 100^\circ$ and 110° . The open and filled triangles correspond to σ^M values for the C8 MSA tensors calculated for energy optima (C2'-endo, $\chi = 68.8^\circ$ and C3'-endo, $\chi = 72.8^\circ$) with the Iglo II and Iglo III basis sets, respectively

differences between the $|\sigma^M|$ values of CSA tensors calculated with the Iglo II and Iglo III basis sets range between 4.6 and 4.9 ppm. For the energy optimum in the *syn* region, the absolute difference between $|\sigma^M|$ values calculated with Iglo II and Iglo III were 4.62 and 4.61 ppm for the C2'-endo ($\chi = 68.8^\circ$) and C3'-endo ($\chi = 72.8^\circ$) sugar pucker, respectively. The effect of atomic basis on the magnitude of C8 MSA tensors of $\sim 4.8 \pm 0.2$ ppm remains essentially constant along the geometry variation carried out for the structural descriptor χ and, furthermore, is independent of sugar pucker.

Discussion

Correlation between experimental and calculated shielding values

In 2006, base carbon CSA values obtained by liquid-crystal NMR and solution relaxation measurements were reported for a double helical A-form RNA segment and a double helical B-form DNA dodecamer (Ying et al. 2006b). These CSA values differ notably from the previous values obtained by solid-state NMR (Stueber and Grant 2002) with changes up to 30 ppm. Table 1 contains a comparison of calculated and experimental $|\sigma^M|$ and η values for base carbons in 2'-deoxyribonucleosides/ribonucleosides. In general, the MSA values calculated in this study were found to be in very good agreement with solid-state NMR results on mono-nucleosides. While calculated $|\sigma^M|$ values differ by an average of about 4.5 ± 4.3 ppm from experimental $|\sigma^M|$ values as determined by solid-state measurements, the corresponding differences between the calculated values and those obtained by liquid-crystal NMR and solution relaxation measurements are approximately 7.9 ± 4.5 ppm. The largest difference is observed for C2 in 2'-deoxyadenosine (up to 15 ppm). However, this difference primarily reflects the fact that our calculations, similarly to the solid-state NMR measurements, do not include the effects of base-pairing. It is well known that the change of the isotropic chemical shift of C2 between base-paired and free adenosine is up to 7 ppm. It follows that the solution values for C2 by Ying et al. (2006b) and those from solid-state by Stueber and Grant and our calculations cannot be directly compared. Considering that the calculated $|\sigma^M|$ values for base carbons are systematically underestimated by approximately 4 ppm (see "Results") due to application of the Iglo II basis in the calculations, we find a very good agreement between the calculated and solution data obtained by liquid-crystal NMR on an A-form RNA and a B-form DNA fragment. It is important to emphasize that while the differences between CSA/MSA values obtained using different approaches are usually less

Table 1 Comparison of $|\sigma^M|$ and η values for nucleoside base carbon CS/MS tensors determined by: quantum chemical calculations (DFT), solid state NMR spectroscopy (SS) by Stueber and Grant (2002),

NMR relaxation and liquid crystal measurements (REL/LC) by Ying et al. (2006b) and liquid crystal measurements (RCSA) by Hansen and Al-Hashimi (2006)

Atom	$ \sigma^M ^a$ (DFT)	$ \sigma^M $ (SS)	$ \sigma^M $ (REL/LC)	$ \sigma^M $ (RCSA)	η^a (DFT)	η (SS)	η (REL/LC)	η (RCSA)
dA-C2	153.6	150*	168 ± 2	n.d.	0.99	0.92*	0.70 ± 0.03	n.d.
dC-C5	140.4	138*	144 ± 1	172.6 ± 21.2*	1.43	1.03*	0.95 ± 0.03	1.40*
dC-C6	192.1	179*	186 ± 3	n.d.	0.86	0.83*	0.67 ± 0.03	n.d.
dT-C6	173.3	170	168 ± 3	n.d.	1.27	1.17	1.02 ± 0.13	n.d.
dG-C8	127.7	126* 134*	133 ± 1	148.1 ± 12.8*	1.4	0.92* 1.08*	0.88 ± 0.05	2.22*
dA-C8	131.1	134*	144 ± 1	n.d.	1.2	1.04*	0.88 ± 0.05	n.d.

The values of $|\sigma^M|$ are given in [ppm]

* The data were derived for ribonucleosides or polyribonucleotides

^a Calculated values obtained in this study taken from 2'-deoxynucleosides with C2'-endo sugar pucker and $\chi \sim 240^\circ$

than 10 ppm, for some base carbons they may be 1.5–2 times larger due to their dependence on orientation of the glycosidic torsion angle χ .

While there is very good agreement among experimental CSA and calculated MSA magnitudes, the asymmetry values for the same CS/MS tensors are mostly very different. The definitions of the CS/MS tensor magnitude and asymmetry imply their different sensitivity towards variations in principal values of the CS/MS-tensor. In general, the asymmetry parameter is much more sensitive to variations of the CS/MS tensor eigenvalues as compared with the magnitude. For example, the magnitude of C6 CS tensor from 2'-deoxythymidine, as determined from solid-state NMR measurements, is 170 ppm, as compared with 168 ppm determined from relaxation measurements. The asymmetry from the solid-state is 1.17, as compared with 1.02 from measurements in solution. In this case, the difference in CS-tensor magnitudes is less than 1.2%, but the difference in the CS-tensor asymmetries is about 12.8%. This example illustrates that even small differences between two CS/MS tensor's eigenvalues might translate into large differences in calculated asymmetries. This makes the comparison between CSA tensors from various NMR experiments and/or MSA tensors from quantum chemical calculations by means of asymmetries problematic.

With the exception of cytidine, Ying et al. (2006b) observed that CSA values of base carbons for 2'-deoxyribonucleosides are essentially the same as for ribonucleosides. The similarity between $|\sigma^M|$ values for 2'-deoxyribonucleosides and ribonucleosides suggest that the base carbons in DNA and RNA might exhibit similar patterns of conformational dependence on sugar pucker and glycosidic torsion angle.

Recently, two independently developed methods suggested RCSAs resulting from weak alignment as a new long-range orientational constraint for NMR refinement of nucleic acid structure (Grishaev et al. 2006; Hansen and

Al-Hashimi 2006). Both of these methods, Hansen's and Grishaev's, express the modulation of the absolute RCSA values as a product of the principal components of the CSA tensor and a simple geometric term relating the orientation of the principal axis of an alignment tensor and the principal axis of the CSA tensor of C2 and C8 purine and C5 and C6 pyrimidine atoms, respectively.

The fundamental assumption of both methods is that the base carbon MSA tensors are independent of molecular geometry. Since the dependence of base carbon MSA tensors on their local environment had not been previously investigated, the authors have suggested avoiding the application of these methods outside the canonical double-helical geometry as it might lead to structural errors (Grishaev et al. 2006). Indeed, our calculations show that the MSA tensor magnitudes and asymmetries of the pyrimidine C6 and purine C8 atoms are strongly modulated by orientation of the glycosidic torsion angle (Figs. 4 and 5), in contrast to the MSA tensor orientations. The differences in the MSA tensor magnitudes between the *syn* and *anti* conformations range from 15 to 25 ppm depending on the base type. Therefore, the method of Grishaev can lead to bias in cases where no a priori information about the conformation of the glycosidic torsion angle is available. Our calculations suggest that for *syn* nucleotides the use of the method might require adjustments of the principal values of the experimental carbon CSA tensors, which seem to be specific for *anti* nucleotides (supplementary material). While application of the Hansen and Grishaev methods to C6 and C8 carbons for *syn* nucleotides may require adjustment of CSA values, our calculation suggests that the application of these methods to quaternary carbons, C4 and C5 in purines and C4 in pyrimidines, should be straightforward. The magnitudes, asymmetries, and orientations of these quaternary base carbon MSA tensors are practically independent of molecular geometry and are not likely to be affected by base-pairing, in contrast to C2

MSA in pyrimidines and purines. With the recent progress in the detection of quarternary carbons in nucleic acids (Fiala et al. 2004; Fiala and Sklenar 2007), quarternary carbon RCSAs may become a valuable source of structural data.

In addition, our calculations suggest that the MSAs of C6 and C8 might be very useful in determining the glycosidic bond preferred conformation due to their strong dependency on χ . For example, simple measurements of cross-correlated relaxation rates between C6/8 CSA and C6/8–H6/8 dipole–dipole should provide unambiguous discrimination between the *syn* and *anti* nucleotides.

Table 2 shows comparison of $|\sigma^M|$ and η values for sugar carbon MSA tensors calculated for *anti* 2'-deoxynucleosides with experimental $|\sigma^M|$ and η values obtained from solid-state measurements on isolated cytidine, guanosine dihydrate, adenosine and 2'-deoxythymidine (Stueber and Grant 2002) or liquid-crystal NMR of A-form

Table 2 Comparison of $|\sigma^M|$ and η values for nucleoside sugar carbon CS/MS tensors determined by: quantum chemical calculations (DFT), solid state NMR spectroscopy (SS) by Stueber and Grant (2002) and NMR liquid crystal measurements (LC) by Bryce et al. (2005)

		dAde/Ade		dGua/Gua		dCyt/Cyt		dThy/Thy	
		$ \sigma^M $	η	$ \sigma^M $	η	$ \sigma^M $	η	$ \sigma^M $	H
C1'	DFT ^a	70.1	1.01	68.3	0.98	69.3	1.04	69.7	1.04
	DFT ^b	55.7	0.53	56.0	0.55	55.5	0.57	56.6	0.60
	SS	28.6*	1.04*	50.6*	0.74*	31.2*	1.31*	65.0	0.88
	LC	30.2*	1.11*						
C2'	DFT ^a	34.6	0.59	34.6	0.64	34.2	0.65	34.1	0.75
	DFT ^b	49.0	1.03	49.8	1.04	48.3	1.03	49.7	1.04
	SS	54.0*	0.76*	52.0*	0.91*	26.1*	0.68*	33.8	1.03
	LC	23.9*	0.85*						
C3'	DFT ^a	33.6	0.57	33.5	0.55	37.6	0.70	37.4	0.73
	DFT ^b	74.1	0.87	74.4	0.89	74.7	0.86	74.6	0.87
	SS	25.5*	0.63*	31.4*	0.72*	58.3*	0.68*	32.4	0.43
	LC	83.1*	0.84*						
C4'	DFT ^a	39.9	0.76	40.5	0.76	45.6	0.64	45.4	0.62
	DFT ^b	51.3	0.66	51.8	0.67	52.2	0.69	52.9	0.70
	SS	57.7*	0.55*	42.3*	0.60*	61.8*	0.52*	47.7	0.88
	LC	83.5*	1.08*						
C5'	DFT ^a	43.1	0.20	42.4	0.19	42.8	0.22	43.0	0.22
	DFT ^b	58.0	0.20	57.5	0.20	58.1	0.18	58.3	0.19
	SS	45.7*	0.63*	39.1*	0.80*	48.5*	0.65*	45.0	0.60
	LC	57.2*	0.23*						

The values of $|\sigma^M|$ are given in [ppm]

* The data were derived for ribonucleosides or polyribonucleotides

^a Calculated values obtained in this study taken from 2'-deoxynucleosides with C2'-endo sugar pucker and $\chi \sim 220^\circ$

^b Calculated values obtained in this study taken from 2'-deoxynucleosides with C3'-endo sugar pucker and $\chi \sim 220^\circ$

helical RNA (Bryce et al. 2005). The data by Stueber and Grant shows differences in sugar CSA tensors among individual nucleosides. However, as the molecular geometries of the nucleosides used for the measurements were unknown and the samples of individual nucleosides differed in preparation, it is not fully clear whether these differences reflect different base types, electrostatic crystal potential, influence of different counter ions or differences in molecular structure. In contrast to results by Stueber and Grant, a fundamental assumption in data analysis by Bryce et al. was that, due to the well-characterized uniformity of A-form RNA, all sugar carbons from the double helical A-type stem might be described by a single CSA tensor regardless of base type. Our calculations indicate that this assumption is also valid for 2'-deoxynucleosides in *anti* conformation. However, there are significant differences between purine and pyrimidine 2'-deoxynucleosides in the *syn* region (Figs. 6 and 7). For C1', the differences between *syn* 2'-deoxypurines and 2'-deoxypyrimidines may be up to 15 ppm for C3'-endo and 10 ppm for C2'-endo sugar pucker. For C3', the differences are even more pronounced, up to 20 ppm. This dependence of C1' and C3' MSAs in 2'-deoxypurines on torsion χ might adversely impact the sugar pucker analysis (Boisbouvier et al. 2000) (*vide infra*).

Pronounced differences in $|\sigma^M|$ values from the RCSA analysis compared with the other methods (DFT, SS, and REL/LC) are observed (Table 1). As discussed in the original report on RCSA analysis by Hansen and Al-Hashimi, these differences might represent an artifact of the RCSA procedure that follows uniform CSA values site to site, bond lengths $r_{CH}(\text{base}) = 1.08 \text{ \AA}$ and $r_{C1'-H1'}(\text{sugar}) = 1.09 \text{ \AA}$, and that internal motions uniformly scale all of the measured residual dipolar couplings (RDCs) and RCSAs by a similar amount. As none of these assumptions is entirely valid for polyribonucleotides, the overestimated $|\sigma^M|$ values as compared with solid-state measurements by Stueber and Grant, relaxation measurements by Ying et al., or the calculated values in this study are most probably indicative of violations of the applied assumptions in the interpretation of RCSA data by Hansen and Al-Hashimi. For example, zero-point motion average bond lengths are substantially larger ($r_{CH}(\text{base}) = 1.102 \text{ \AA}$ and $r_{C1'-H1'}(\text{sugar}) = 1.118 \text{ \AA}$) than those used by Hansen and Al-Hashimi. The use of zero-point motion average bond lengths would result in reduced values of $|\sigma^M|$. On the other hand, these differences might also reflect genuine differences in aromatic carbon electron densities in mononucleosides and polynucleosides.

The only available experimental data on sugar carbon CSA in 2'-deoxynucleosides are for 2'-deoxythymidine (Stueber and Grant 2002). The calculated $|\sigma^M|$ values are in excellent agreement with the experimental results of Stueber and Grant. The calculated $|\sigma^M|$ values differ from

the previously reported experimental values by an average of 2.9 ± 2 ppm. Remarkably, the variations in sugar carbon MSA magnitudes, due to sugar repuckering, range between 5 and 35 ppm.

Not surprisingly, the sugar carbon MSA tensors for 2'-deoxynucleosides differ in absolute numbers from those in ribonucleosides (Table 2). However, it may be expected that sugar carbon MSA tensors in (poly)-ribonucleotides will exhibit similar patterns of conformational dependence on sugar pucker and glycosidic torsion angle. To illustrate this, we compared experimentally acquired polyribonucleotide λ values, defined as $\lambda = \frac{I_{C3',C3'-H3'}^{CSA,DD}}{I_{C1',C1'-H1'}^{CSA,DD}} \cong \frac{\sigma^*(C3')}{\sigma^*(C1')}$, where

$\sigma^* = \sum_{i=1}^3 \sigma_{ii}^C \left(\frac{3 \cos^2 \theta_{ii} - 1}{2} \right)$, σ_{ii}^C is the ii -th component of the diagonalized $C1'/C3'$ CSA/MSA tensor, and θ_{ii} is the projection angle between $C1'-H1'/C3'-H3'$ dipole-dipole vector and principal axis of the CSA/MSA tensor, from Boisbouvier et al. (2000) with the λ values predicted by our calculations (Fig. 9). In their study, Boisbouvier et al. measured cross-correlated relaxation rates between $C1'-H1'/C3'-H3'$ dipole-dipole and $C1'/C3'$ chemical shift anisotropy in three model RNA polynucleotides.

They demonstrated that the ratio λ sensitively reflects differences in magnitude and orientation of the $C1'/C3'$ MS tensors between $C2'$ -endo and $C3'$ -endo sugar puckers, as originally suggested by Dejaegere and Case (1998). The sugar pucker in polyribonucleotides can be unequivocally assigned based on the λ values from calculated MSA tensors for 2'-deoxynucleosides (Fig. 9). This indicates that the differences in magnetic shielding due to sugar

repuckering override those due to differences in the molecular structure of ribose and 2'-deoxyribose.

In their original report, Boisbouvier et al. assigned $\lambda < 1.0$ and $\lambda > 2.0$ to $C3'$ -endo and $C2'$ -endo sugar conformations, respectively. Values of λ around 1.5 were assigned to sugars undergoing conformational averaging. For 2'-deoxyguanosine, our calculations suggest that λ values around 1.5 might alternatively be interpreted as indicating *syn* 2'-deoxyguanosine with $C3'$ -endo sugar pucker.

In a recent report, Ravindranathan et al. (2005) characterized base carbon CSA tensors in an RNA kissing complex using both transverse and longitudinal auto- and cross-correlated relaxation rates. They found that the C2 and C8 CSA tensor magnitudes for residues in a stem differ from those in a loop region. Differences in $|\sigma^M|$ of up to 14.2 ppm for adenosine and 19.4 ppm for guanosine residues were observed. However, these differences could not be attributed directly to either reorientation of the glycosidic bond (all residues in the loop as well as the stem are in the *anti* conformation) or internal dynamics. These results indicate that, in addition to the conformationally dependent variability of the MSA tensors and internal dynamics effects, there are other factors influencing actual MSA values. Extra care must be taken when interpreting MSA-related NMR parameters in the non-canonical region, as minute changes in hydration and/or accessibility of ion binding sites might also impact chemical shielding.

It follows that despite the good agreement of our calculated MSA values with experimental data, casual application of our calculated MSA values can still lead to biased assessment of the structural information. The accuracy of the calculated MSA tensor depends on the model compound and basis set used in the calculations. In order to assess MSA values correctly, the calculated points should be obtained from polynucleotide models in the explicit solvent and in the presence of ions. In practice, however, it is very difficult to satisfy such a requirement due to exorbitant computational time costs. Hence, evaluation of NMR observables based on calculated MSA values must be done with caution. On the other hand, the constant and systematic effect of the atomic basis on $|\sigma^M|$ is less important for the prediction of its actual structural dependence since experimentalists are usually interested in relatively distinguishing between different conformers. The absolute correlation of experimental and theoretical values is the domain of methodological studies.

While finishing our calculations, we provided our preliminary data to the group of Prof. Harald Schwalbe for analysis of the temperature-dependent dynamics of RNA YNMG-tetraloops. Schwalbe and coworkers found that the χ -value dependence of the MSA needs to be considered in order to yield fully consistent results on order parameters

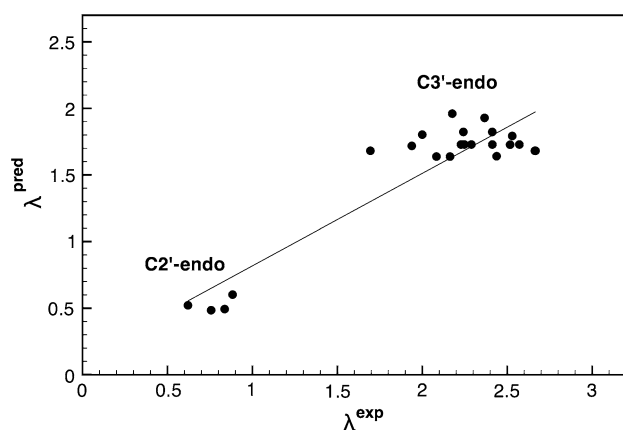


Fig. 9 Correlation plot between experimental $\exp \lambda = \frac{I_{C3',C3'-H3'}^{CSA,DD}}{I_{C1',C1'-H1'}^{CSA,DD}} \cong \frac{\sigma^*(C3')}{\sigma^*(C1')}$ values acquired on three different RNA molecules by Boisbouvier et al. (2000) and those calculated based on $C1'$ and $C3'$ MSA tensors derived in present study (λ^{pred}). $\sigma^* = \sum_{i=1}^3 \sigma_{ii}^C \left(\frac{3 \cos^2 \theta_{ii} - 1}{2} \right)$, where σ_{ii}^C is the ii -th component of the diagonalized $C1'/C3'$ MSA tensor, and θ_{ii} is the projection angle between $C1'-H1'/C3'-H3'$ dipole-dipole vector and principal axis of the MSA tensor

derived from ^{13}C relaxation rates or ^{15}N relaxation analysis (Ferner et al. 2008). Their results have clearly demonstrated that the conformation around the glycosidic torsion angle has a pronounced effect on the base carbon MSA that propagates into the order parameter analysis. Importantly, they showed that calculated MSA values, such as those generated here, can be used directly for correction of experimental ^{13}C and ^{15}N CSA values, facilitating an appropriate and physically meaningful model-free analysis.

Strictly speaking, the presented calculated MS-tensors parameters are approximate. A more quantitative analysis would have to respect dynamic effects, i.e., quantum-mechanical averaging of the MS-tensors over the bond lengths and bond angle fluctuations and torsional motions. So far, there are no data available on the effect of vibrational averaging on chemical shielding anisotropies in nucleic acids. However, Jordan et al. recently showed that the principal values and orientations of the ^{13}C carbonyl MSA tensor in peptides are very sensitive to small local changes in structure. In analogy to proteins, one might expect that the MSA tensors in nucleic acid will fluctuate as a function of time when bond lengths and bond and torsion angles fluctuate. Based on recent calculations by Jordan et al. (2007) and Tang and Case (2007), it is possible to estimate that the magnitude of static MSA tensors would be about 5–15% higher as compared with a vibrationally averaged, effective CS tensor derived from measurements in solution.

We showed that the use of the Iglo II basis in our calculation leads to systematic underestimation of the static MS tensor. In this respect, the virtual quantitative agreement between our calculated MSA values and those from relaxation and liquid crystal measurements might be considered as an artifact. The agreement between experimental and calculated values rather indicates that underestimation of the static MSA values due to the use of a small basis set in our calculations on average compensates for the neglected averaging effects.

Conclusion

In this study, we investigated the dependence of ^{13}C and ^{15}N MSA tensors in 2'-deoxynucleosides on the conformation of the glycosidic torsion angle and sugar pucker mode using DFT calculations. Contrary to the assumptions generally applied in structural interpretations of NMR parameters related to MSA, such as transversal and longitudinal auto- and cross-correlated relaxation rates and residual chemical shift anisotropies, our calculations indicate that the conformation of the glycosidic bond strongly influences C6/8 MSA tensors in 2'-deoxynucleosides. In light of the methods recently established for structural

interpretation of base carbon RCSA (Grishaev et al. 2006; Hansen and Al-Hashimi 2006), we expect that the dependence of C6 and C8 CSA tensors on χ may adversely affect RCSA analysis and lead to errors in structure determination.

On the other hand, our calculations indicate that this dependence may be very useful for determining preferred glycosidic bond conformations. Based on our calculations, we propose to extend Hansen's and Grishaev's methods to quarternary C4 in 2'-deoxypyrimidines and to C4, C5, C6, and N7 in 2'-deoxypurines. In general, the use of conformation-specific MSA tensor values appears to be crucial for proper interpretation of NMR data related to MSA for both canonical and non-canonical nucleic acid structures. Good agreement of data calculated for 2'-deoxynucleosides with experimental data acquired on ribonucleosides and polyribonucleotides suggests that the conformational dependence of ^{13}C and ^{15}N MSA suggested by our calculations may be valid for polyribonucleotides as well.

Acknowledgements This study was supported by Ministry of Education of the Czech Republic grant no. MSM6007665801 and by the Academy of Sciences of the Czech Republic, grants no. IAA400550701/AV0Z50040507/AV0Z50040702 and KAN200100 801. We thank Prof. H. Schwalbe for providing us with a manuscript (Ferner et al. 2008) prior to publication.

References

- Akke M, Fiala R, Jiang F, Patel D, Palmer AG III (1997) Base dynamics in a UUCG tetraloop RNA hairpin characterized by ^{15}N spin relaxation: correlations with structure and stability. *RNA* 3:702–709
- Altona C, Sundaralingam M (1972) Conformational-analysis of sugar ring in nucleosides and nucleotides—new description using concept of pseudorotation. *J Am Chem Soc* 94:8205–8212
- Becke AD (1993) A new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 98:1372–1377
- Boisbouvier J, Brutscher B, Pardi A, Marion D, Simorre J-P (2000) NMR determination of sugar puckers in nucleic acids from CSA-dipolar cross-correlated relaxation. *J Am Chem Soc* 122:6779–6780
- Bryce DL, Grishaev A, Bax A (2005) Measurement of ribose carbon chemical shift tensors for A-form RNA by liquid crystal NMR spectroscopy. *J Am Chem Soc* 127:7387–7396
- Czernek J, Fiala R, Sklenar V (2000) Hydrogen bonding effects on the (^{15}N) and (^1H) shielding tensors in nucleic acid base pairs. *J Magn Reson* 145:142–146
- Dejaegere AP, Case DA (1998) Density functional study of ribose and deoxyribose chemical shifts. *J Phys Chem A* 102:5280–5289
- Duchardt E, Schwalbe H (2005) Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR ^{13}C relaxation. *J Biomol NMR* 32:295–308
- Duchardt E, Richter C, Ohlenschlager O, Gorchach M, Wohnert J, Schwalbe H (2004) Determination of the glycosidic bond angle χ in RNA from cross-correlated relaxation of CH dipolar coupling and N chemical shift anisotropy. *J Am Chem Soc* 126:1962–1970

- Ebrahimi M, Rossi P, Rogers C, Harbison GS (2001) Dependence of ^{13}C NMR chemical shifts on conformations of RNA nucleosides and nucleotides. *J Magn Reson* 150:1–9
- Ferner J, Villa A, Duchardt E, Widjajakusuma E, Wohnert J, Stock G, Schwalbe H (2008) NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops. *Nucleic Acids Res* 36:1928–1940
- Fiala R, Sklenar V (2007) ^{13}C -detected NMR experiments for measuring chemical shifts and coupling constants in nucleic acid bases. *J Biomol NMR* 39:153–163
- Fiala R, Munzarova ML, Sklenar V (2004) Experiments for correlating quaternary carbons in RNA bases. *J Biomol NMR* 29:477–490
- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03, revision C.02, Gaussian Inc., Pittsburgh, PA
- Grishaev A, Ying J, Bax A (2006) Pseudo-CSA restraints for NMR refinement of nucleic acid structure. *J Am Chem Soc* 128:10010–10011
- Hansen AL, Al-Hashimi HM (2006) Insight into the CSA tensors of nucleobase carbons in RNA polynucleotides from solution measurements of residual CSA: towards new long-range orientational constraints. *J Magn Reson* 179:299–307
- Hu J-Z, Facelli JC, Alderman DW, Pugmire RJ, Grant DM (1998) ^{15}N chemical shift tensors in nucleic acid bases. *J Am Chem Soc* 120:9863–9869
- Jordan DM, Mills KM, Andricioaei I, Bhattacharya A, Palmo K, Zuiderweg ER (2007) Parameterization of peptide ^{13}C carbonyl chemical shielding anisotropy in molecular dynamics simulations. *ChemPhysChem* 8:1375–1385
- Kutzelnigg W, Fleischer U, Schindler M (1991) The IGLO Method: ab initio calculation and interpretation of NMR chemical shifts and magnetic susceptibilities. In: Diehl P, Fluck E, Günther H, Kosfeld R, Seelig J (eds) *NMR basic principles and progress*, vol 23. Springer, Heidelberg, pp 165–262
- Lee CT, Yang WT, Parr RG (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys Rev B* 37:785–789
- Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wuthrich K (1998) Recommendations for the presentation of NMR structures of proteins and nucleic acids. *J Mol Biol* 28:933–952
- Precechtelova J, Munzarova ML, Novak P, Sklenar V (2007) Relationships between ^{31}P chemical shift tensors and conformation of nucleic acid backbone: a DFT study. *J Phys Chem B* 111:2658–2667
- Ravindranathan S, Kim CH, Bodenhausen G (2003) Cross correlations between ^{13}C - ^1H dipolar interactions and ^{15}N chemical shift anisotropy in nucleic acids. *J Biomol NMR* 27:365–375
- Ravindranathan S, Kim CH, Bodenhausen G (2005) Determination of ^{13}C CSA tensors: extension of the model-independent approach to an RNA kissing complex undergoing anisotropic rotational diffusion in solution. *J Biomol NMR* 33:163–174
- Schofberger W, Sychrovsky V, Trantirek L (2006) Quantum chemical calculations of NMR parameters. In: Sponer J, Lankas F (eds) *Computational studies of RNA and DNA (Challenges and Advances in Computational Chemistry and Physics)*. Springer, Dordrecht, pp 513–536
- Shajani Z, Varani G (2007) NMR studies of dynamics in RNA and DNA by ^{13}C relaxation. *Biopolymers* 86:348–359
- Sitkoff D, Case DA (1998) Theories of chemical shift anisotropies in proteins and nucleic acids. *Prog NMR Spectr* 32:165–190
- Stueber D, Grant DM (2002) ^{13}C and (^{15}N) chemical shift tensors in adenosine, guanosine dihydrate, 2'-deoxythymidine, and cytidine. *J Am Chem Soc* 124:10539–10551
- Sychrovsky V, Muller N, Schneider B, Smrecki V, Spirko V, Sponer J, Trantirek L (2005) Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J Am Chem Soc* 127:14663–14667
- Tang S, Case DA (2007) Vibrational averaging of chemical shift anisotropies in model peptides. *J Biomol NMR* 38:255–266
- Trantirek L, Caha E, Kaderavek P, Fiala R (2007) NMR (^{13}C)-relaxation study of base and sugar dynamics in GCAA RNA hairpin tetraloop. *J Biomol Struct Dyn* 25:243–252
- Wolinski K, Hinton JF, Pulay P (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical-shift calculations. *J Am Chem Soc* 112:8251–8260
- Ying J, Grishaev A, Bax A (2006a) Carbon- 13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. *Magn Reson Chem* 44:302–310
- Ying J, Grishaev A, Bryce DL, Bax A (2006b) Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. *J Am Chem Soc* 128:11443–11454