The Combined Treatment of the Intrinsic and External Contributions to the Stability of the Nucleic Acids Containing Unnatural Nucleosides

JIRÍ CZERNEK
Department of Bioanalogous and Special Polymers
Institute of Macromolecular Chemistry
Heyrovsky Square 2, 16206 Prague
THE CZECH REPUBLIC

Abstract: The intrinsic and external factors governing the stability of unnatural motifs in nucleic acids are discussed. The former are described by means of high-level ab initio calculations on the investigated base pairs, while the latter are derived from an analysis of the molecular dynamics trajectories. Their combined description is proposed, and a short DNA duplex containing the 3-fluorobenzene self-pair is treated as the test case for the new approach.

Key-Words: nucleic acid, nucleobase, hydrogen-bonding, stacking, solvation, molecular dynamics, ab initio

1 Introduction
An incorporation into the nucleic acids (i.e., DNA or RNA) of the compounds other than the classical nucleobases (adenine, A, cytosine, C, guanine, G, thymine, T, and uracil, U) is a vital area of current biophysical and biochemical research. It covers a broad range of highly important topics comprising but not limited to, in the case of the modified DNA, an expansion of the genetic alphabet [1]; the studies on the replication [2]; development of novel tools for the detection of biomolecular structures and events [3]. For RNA, the investigations are mainly aimed at achieving new functionality as ligands [4] or catalysts [5] of the modified RNAs obtained after the transcription involving the unnatural base pairs [6].

Previous efforts to create alternative base pairs for replication and transcription have relied on nonstandard hydrogen-bonding schemes that differ from those of the Watson–Crick (WC) base pairs (i.e., G – C, A – T in DNA and A – U in RNA) [7]. Unfortunately, their practical usage was found to be hampered by misincorporations resulting in noncognate pairing with classical bases. Consequently, the new type of unnatural base pairs, in which pairing is mediated by hydrophobicity and complementarity of shape, has been devised [8] and extensively tested for replication efficiency and fidelity (see ref. [9] for a recent review). The aromatic surface area and heteroatom substitution pattern of the nucleobase surrogates have been found to play the decisive role for the polymerase recognition and efficient replication of the modified nucleic acids, and the structure–activity relationship(s) have been sought. The interplay between those two factors can be complex [10], which is further complicated by a possibility of the application of either self-pairs or heteropairs of unnatural nucleobases. Nonetheless, the theories of molecular recognition can be applied to the design and development of unnatural base pairs [6], [11].

Undoubtedly, molecular modeling can be usefully employed in the rational design of the unnatural motifs in nucleic acids (see ref. [12] for a recent example). The underlying assumption is the additive separation of the total stabilization provided by a given structural element into its intrinsic and external components. This concept has been successfully applied to nucleic acids, with the stabilization energy of the base pairs being the intrinsic contribution, while the role of environment in the duplex functioning is covered by the external component (cf. ref. [13]). Thus, the intrinsic contribution can be accurately estimated from high-level ab initio calculations. In particular, the second-order Møller–Plesset perturbation theory (MP2) [14] supermolecular interaction energy can be extrapolated to the complete basis set (CBS) limit [15], and the CCSD(T) (the singles and doubles coupled cluster method with noniterative inclusion of triple electron excitations [15]) correction term is added to it to arrive at the “CCSD(T)/CBS” value, which is usually of chemical accuracy (within ±one kcal/mol from the experiment). Recently an important alternative for an estimation of the base pairs’ interaction energies emerged, i.e., the computationally efficient variant of the symmetry-adapted perturbation theory (cf. ref. [16]).

The estimation of the external contribution is less straightforward due to the delicate balance of the hydrogen-bonding, stacking, solvation and other effects, which can be monitored by the molecular dynamics (MD) simulations [17]. Here we propose to analyze the MD trajectories in order to discern the external effects by assuming that the unnatural base pair moves in the effective field exerted by the environment (i.e., the rest
of the duplex, solvent molecules, the counterions), which
determines the overall stabilization as the time-average
of the corresponding interaction energy. The latter can
be reliably obtained from the CCSD(T)/CBS
calculations mentioned above. This combined approach
is applied here to a short DNA duplex to study the
influence of the 3-fluorobenzene self-pair (see below)
incorporated into it.

2 Methods

2.1 Molecular dynamics simulations
Romesberg and his coworkers [18] have described the
favorable replication properties of the 3FB self-pair
(3FB stands for the 3-fluorobenzene nucleobase analogue).
Consequently, they have synthesized the
oligonucleotides 5'-d(GCAGA3FBCGATCC) and 5'-
d(GGATCG3FBTCTGC) and characterized the resulting
duplex by circular dichroism (CD) and nuclear magnetic
resonance (NMR) spectroscopical techniques. The
structure and dynamics of this duplex has been studied
here by means of the MD simulations. All calculations
have been performed using the Amber 7 package [19]
(cf. its manual [20] for references to the techniques
described in the rest of Part 2.1). Thus, to comply with
the spectroscopical evidence [18], the starting structure
has been built in the canonical B-form using the nucgen
module, and is shown in Figure 1. It has been immersed
in the periodic box consisting of around 21 thousands
TIP3P water molecules (initial box size 60 Å × 40 Å ×
40 Å) and 22 sodium cations, which were used to
neutralize the phosphate backbone charge. Simulations
have been carried out using the sander module in the
isothermic isobaric ensemble (p = 1 Atm, T = 298 K).
The parameters for 3FB and its nucleoside, which are
missing in the default force field, have been obtained by
standard techniques, i.e., employing the ab initio
Hartree–Fock (HF) calculations performed with the
6-31G* basis set and followed by fitting procedures to
estimate the new force constants and partial charges.
Periodic boundary conditions and the Particle Mesh
Ewald algorithm have been adopted. All bond lengths
have been frozen using SHAKE, thus allowing a 2 fs
integration time step. At first the standard warm-up
procedures have been applied (165 ps in total) to a)
remove the strain in the initial structure; b) equilibrate
the water molecules and counterions around the solute;
c) gradually remove the restraints on the duplex in the
presence of solvent and sodium ions while increasing the
temperature. They have been followed by unrestrained
20 ns production runs, which have been repeated thrice.
Snapshots have been saved every 10 ps and averaged
over the three trajectories for subsequent analysis.

Fig 1. The initial structure of the duplex (the 3-
fluorobenzene self-pair is in the middle, with the fluorine
atoms shown in yellow).

2.2 The target geometries
The geometries of the 3FB self-pair as provided in the
course of the MD simulations of the duplex were treated
as follows (cf. Fig. 2).
First, the fluorobenzene structure from the
oligonucleotide 5'-d(GCAGA3FBCGATCC) (further
referred to as S1) has been positioned so that the
fluorine-bearing carbon atom C1 lies in the origin of the
coordinate system and the coordinates of the fluorine
atom F1 are (rCF, 0, 0), where rCF denotes the length of
the C1–F1 bond.

Fig 2. The coordinate system for the translations and
rotations described in the text (the z-axis is perpendicular
to the page).
Second, using simple trigonometry, the translations \( \{V, W\} \) and rotations \( \{R_\alpha, R_\beta, R_\gamma\} \) of the fluorobenzene structure from the oligonucleotide \( 5'- \text{d(GGATCG} \text{3FBCTG)C} \) (further referred to as \( \text{S2} \)) have been derived, which would position \( \text{S2} \) so that the \( F^2 \) atom lies in the origin of the coordinate system and the coordinates of the fluorine-bearing carbon atom \( C_2 \) are \( (r_{CF}, 0, 0) \). Thus, the translation \( V \) is the center of mass separation between \( \text{S1} \) and \( \text{S2} \), and \( W \) moves the \( \text{S2} \) structure in the \( xy \)-plane. The rigid body rotations \( R_\alpha, R_\beta, \) and \( R_\gamma \), are described by the rotation matrices given by Equations (1) – (3), with the angles \( \alpha, \beta \) and \( \gamma \) representing the left-hand sense rotations in the \( xy, xz \) and \( yz \) planes, respectively. Thus, an application of the translations and rotations specified above defines the spatial arrangement of \( \text{S2} \) with respect to \( \text{S1} \), and it has been extracted by averaging over the MD snapshots.

\[
R_\alpha = \begin{pmatrix}
\cos(\alpha) & -\sin(\alpha) & 0 \\
\sin(\alpha) & \cos(\alpha) & 0 \\
0 & 0 & 1
\end{pmatrix} \\
R_\beta = \begin{pmatrix}
\cos(\beta) & 0 & \sin(\beta) \\
0 & 1 & 0 \\
-\sin(\beta) & 0 & \cos(\beta)
\end{pmatrix} \\
R_\gamma = \begin{pmatrix}
1 & 0 & 0 \\
0 & \cos(\gamma) & -\sin(\gamma) \\
0 & \sin(\gamma) & \cos(\gamma)
\end{pmatrix}
\]

It should be noted that both \( \text{S1} \) and \( \text{S2} \) have been assumed to be planar, and their reference geometries have been taken from the fully optimized MP2/aug-cc-pVDZ structure of an isolated fluorobenzene molecule obtained using default algorithms and settings of Gaussian 03 suite of quantum chemical programs [21].

### 2.3 Ab initio calculations

The counterpoise-corrected [22] CCSD(T)/CBS interaction energies \( \Delta E \) of the fluorobenzene arrangements have been obtained using the variational supermolecular method by combining two approaches. The first one applies the resolution of the identity (RI) approximation for the calculation of the MP2 energies [23], and adopts the family of the augmented correlation-consistent polarized-valence \( X \)-tuple basis sets, aug-cc-pVXZ, where \( X \) is the cardinal number associated with each basis set [24]. Thus, the RI-MP2 energies have been calculated for the cardinal numbers \( X = 2 \) (double-\( \xi \), DZ), 3 (TZ) and 4 (QZ) and subsequently the CBS energy estimates \( E_{\text{CBS}} \) have been obtained from the mixed Gaussian/exponential form as given by Equation (4)

\[
E(X) = E_{\text{CBS}} + b \exp[-(X-1)] + c \exp[-(X-1)^2]
\]

The \( E_{\text{CBS}} \) data have been then employed to get the RI-MP2 CBS supramolecular interaction energies \( \Delta E_{\text{CBS}} \).

The second approach employs the aug-cc-pVdz basis set and calculates the difference between the corresponding MP2 and CCSD(T) interaction energies to approximate the higher-order contributions to the correlation energy [13]. Resulting correction term to the interaction energy \( \Delta E_{\text{corr}} \) has been added to the \( \Delta E_{\text{CBS}} \) value to obtain the \( \Delta E \), an estimate of the intrinsic contribution of a given base pair to the overall stability of the DNA duplex. The MP2 and CCSD(T), and RI-MP2, calculations have been performed, accordingly, using Molpro 2006.1 [25] and Turbomole V5-7-1 [26] program packages.

### 3 Results and Discussion

It has been established by hard-core quantum chemical calculations that the global minimum of the fluorobenzene dimer corresponds to the \( C_{2h} \)-symmetric structure [27]. Here its geometry has been recomputed at the MP2/aug-cc-pVDZ level to obtain the reference data for further analyses. Thus, the displacements defined above have been found to be \( V = 6.58 \) and \( W = 2.13 \) \( \text{Å} \), and the rotational angles are obviously zero due to the structure’s symmetry. The \( \Delta E_{\text{CBS}} \) and \( \Delta E_{\text{corr}} \) contributions are 10.76 and 0.12 kJ/mol, giving the \( \Delta E \) of 10.89 kJ/mol. An error stemming from the application of the rigid monomers, instead of the relaxed dimer, in the calculations has been found to be negligible: the \( \Delta E_{\text{CBS}} \) and \( \Delta E_{\text{corr}} \) values are 10.45 and 0.08 kJ/mol, respectively.

Hence, the interaction energies have been evaluated for the arrangement of the \( \text{S1} \) and \( \text{S2} \) (Figure 3), which has been obtained by averaging the values from the MD trajectories as detailed in part 2.2. It is described by the displacements of \( V = 6.89 \) and \( W = 4.35 \) \( \text{Å} \), and the rotation angles \( \alpha = 32^\circ, \beta = 56^\circ, \gamma = 29^\circ \). Notably, the interaction energies do not significantly differ from the values obtained for the global minimum (see above), with \( \Delta E_{\text{CBS}}, \Delta E_{\text{corr}} \) and \( \Delta E \) amounting to 9.69, 0.10 and 9.80 kJ/mol, respectively. Clearly, this finding is consistent with the experimental observation [18] that the \( 3\text{FB} \) unnatural base pair does not significantly perturb the B-DNA duplex.

**Fig 3.** The averaged structure of the fluorobenzene base pair obtained from the MD simulations.
4 Conclusion
A novel approach for the treatment of the intrinsic and external factors governing the stability of unnatural motifs in nucleic acids has been elaborated. It utilizes the MD simulations to obtain the geometries representing the unnatural base pair in the effective field of its physico-chemical environment, and estimates the CCSD(T)/CBS interaction energy. An application of this method to a short DNA fragment containing the 3-fluorobenzene self-pair has indicated that this base pair is not significantly distorted in the duplex, which is consistent with the experimental results [18].

Acknowledgement
This work was supported by the Academy of Sciences of the Czech Republic (Grant KJB400500602). Time allocation in the Czech Academic Supercomputer Centre is gratefully acknowledged.

References:

