Local mode analysis of characteristic vibrational coupling in nucleobases and Watson–Crick base pairs of DNA

To cite this article: Mateus Quintano et al 2022 Electron. Struct. 4 044005

View the article online for updates and enhancements.
Local mode analysis of characteristic vibrational coupling in nucleobases and Watson–Crick base pairs of DNA

Mateus Quintano, Alexis A A Delgado, Renaldo T Moura Jr, Marek Freindorf and Elfi Kraka

Abstract

Two-dimensional infrared spectroscopy has reported highly delocalized in-plane base vibrations in the fingerprint region of nucleotide monophosphates, suggesting the involvement of base pair C=O and C=C ring bonds and considerable interaction between C=O bond stretches. The high delocalization results in congested vibrational spectra, which complicates the assignment of the peaks. This congestion also extends to Watson–Crick base pairs. We applied in this work the characterization of normal mode procedure, a special feature of our local mode analysis, and could for the first time identify the C=O and C=C bonds being engaged in base pair coupling and quantify their contribution to each of the delocalized fingerprint vibration. In addition, a detailed and quantitative description of the hydrogen bonds involved in the Watson–Crick base pairs was provided. Based on the results of this study, we developed a new protocol to elucidate on the assignment of bands in the vibrational spectra of nucleic acids by probing the vibrational space for specific interactions between functional groups prior to and upon base pairing. This protocol will aid to fill the gap between deoxyribonucleic acid structural information and vibrational spectroscopy experiments by facilitating the interpretation of spectra on a quantitative basis.

1. Introduction

Deoxyribonucleic acid (DNA) is a nucleotide-made polymer of genetic, catalytic and structural importance, whose double-stranded nature is governed by complementary base pairing through specific hydrogen bonds [1]. As illustrated in figure 1, the resulting Watson–Crick structure is comprised of two types of base pairs, that is, adenine (A) and guanine (G) must pair with thymine (T) and cytosine (C), respectively [1, 2]. Information on the structure of nucleic acids and their interactions are of particular interest to linear and nonlinear vibrational spectroscopy investigations [3–20]. Within the frequency range of 1500–1800 cm⁻¹ are the in-plane base vibrations (fingerprint or characteristic region).

In computational investigations in-plane base vibrations are usually restricted to the nucleobases with the methyl group as a cost-effective replacement for the phosphate and sugar groups [6, 7]. These simulations usually take into account hydration effects and deuterium in place of labile amino hydrogen atoms to provide a more realistic quantum chemical approach to vibrational spectra [3, 6, 7, 10, 14]. Deuterium effects are accounted for as experimental measurements are usually performed in deuterated water to circumvent interference of the HOH bending mode [19]. Deuterium effects dramatically change the fingerprint region of the spectra by shifting ND and ND₂ scissoring and stretching mode frequencies to regions outside the characteristic one [3, 7]. Hence, in the realm of deuteration effects, the focus is on C=O stretching and ring deformation modes [7].
Figure 1. Suggested C=O and C=C bonds (in color) of methylated nucleobases A, G, T, and C models being involved in base pair coupling [10, 18, 19]. Index numbers refer to the numbering of atoms used in this work. Terminal carbon atoms of the methyl groups are indicated by numbers without atomic symbols.

The sensitivity of IR spectroscopy to changes in the secondary structure of DNA can be linked to specific normal vibrational modes in the fingerprint region [3–7, 15–20]. In this respect, the search for a theoretical framework that contributes to the elucidation of the couplings (interactions) involving C=O and C=C vibrations has been of recent interest [10, 18, 19]. It has been suggested that particular nucleobase C=C and C=O bonds engage in base pair couplings, as highlighted in figure 1. However, the high delocalization of normal vibrational modes has made it hard to pinpoint bond stretching contributions even for the building blocks of nucleic acids, as revealed by 2D IR measurements of the in-plane base vibrations of nucleotide monophosphates [10]. Normal vibrational modes usually tend to extend over aromatic rings, contain interaction between C=O stretches, and exhibit interdependence between the C=O and C=C groups in the pyrimidine bases [10]. Hydrogen bonding in Watson–Crick base pairs has been reported to influence delocalization and coupling of the fingerprint normal vibrational modes as well [4, 6–9]. Clearly, all of this has called for new ideas to facilitate the interpretation of the vibrational spectra of nucleic acids, which we have materialized in this work via the local mode analysis (LMA), originally proposed by Konkoli and Cremer [21–26] using the characterization of normal mode (CNM) procedure [23, 25] as a key tool.

2. Method

2.1. The Wilson GF formalism

The normal mode analysis machinery describing a vibrating molecule with \( N \) atoms is based on the Wilson equation of vibrational spectroscopy [27–31]

\[
\mathbf{F}' \mathbf{L} = \mathbf{M} \Lambda \mathbf{L}^\dagger
\]  

(1)

where \( \mathbf{F}' \) is the \( 3N \times 3N \) force constant matrix (Hessian matrix) in Cartesian coordinates \( x_i \) (\( i = 1, \cdots, 3N \)). The diagonal \( 3N \times 3N \) mass matrix \( \mathbf{M} \) contains the atomic mass for each atom three times to account for the motion in \( x, y \) and \( z \) directions. The diagonal matrix \( \mathbf{A} \) contains the eigenvalues \( \lambda_\mu \) (\( \mu = 1, \cdots, N_{\text{vib}} \); \( N_{\text{vib}} = 3N - \Theta, \Theta = 5 \) or \( 6 \) for linear and non-linear molecules, respectively), whose relationship with the normal vibrational mode frequencies \( \omega_\mu \) is given by \( \lambda_\mu = 4\pi^2c^2\omega_\mu^2 \), with \( c \) being the speed of light. The \( 3N \times N_{\text{vib}} \) matrix \( \mathbf{L} \) collects the normal modes \( l_\mu \) as orthonormal column vectors.

One can also express the molecular geometry in terms of internal coordinates \( \mathbf{q} \) rather than Cartesian coordinates \( \mathbf{x} \), and by this the Wilson equation adopts a new form [27]:

\[
\mathbf{F} \mathbf{D} = \mathbf{G}^{-1} \mathbf{D} \mathbf{A}
\]

(2)
where $\mathbf{D}$ collects the normal mode vectors $\mathbf{d}_n (\mu = 1, \ldots, N_{\text{vib}})$ column-wise, and the Wilson matrix $\mathbf{G}$, which is defined as

$$\mathbf{G} = \mathbf{BM}^{-1}\mathbf{B}^T$$

whose inverse represents the kinetic energy in terms of internal coordinates. The elements of the rectangular $\mathbf{B}$ matrix in equation (3) are defined by the partial derivatives of internal coordinates $q_n (n = 1, 2, 3 \ldots N_{\text{vib}})$ with regard to Cartesian coordinates $x_i (i = 1, 2, 3 \ldots 3N)$,

$$\mathbf{B}_n = \frac{\delta q_n (x)}{\delta x_i}.$$  

It is important to note that the $\mathbf{B}$ matrix plays a central role for the Wilson equation of spectroscopy, namely connecting different sets of coordinates (internal, symmetry, curvilinear, etc) [32–34] or Cremer-Pople ring puckering coordinates [35], with the Cartesian coordinates [27]. Therefore, whenever a new set of coordinates is introduced, the first step is to derive the appropriate $\mathbf{B}$ matrix, as demonstrated below for the definition of intermonomer modes.

Diagonalization of equation (2) leads to a transformation to normal coordinates $\mathbf{Q}$ and the diagonal force constant matrix $\mathbf{K}$ expressed in normal coordinates (i.e. $\mathbf{K} = \mathbf{K}^0$) which both play a central role in the normal mode analysis [27]

$$\mathbf{D}^T \mathbf{F}^0 \mathbf{D} = \mathbf{K}.$$  

The transformation to normal coordinates $\mathbf{Q}$ leading to the diagonal force constant matrix $\mathbf{K}$ and normal mode vectors $\mathbf{d}_n$ (see equation (5)) is a standard procedure in modern quantum chemistry packages calculating vibrational frequencies, providing access to important electronic structure information of a molecule and the motion of its atoms for each of the $N_{\text{vib}}$ vibrations [36–38]. Therefore, vibrational frequencies and related force constants have become a popular measure of bond strength.

However, caveat is appropriate. As already pointed out by Wilson in 1941 [39], normal coordinates $\mathbf{Q}$ are generally a linear combination of internal coordinates $\mathbf{q}$ or Cartesian coordinates $\mathbf{x}$

$$\mathbf{Q} = \tilde{\mathbf{D}}^T \mathbf{G}^{-1} \mathbf{q} = \tilde{\mathbf{D}}^T \mathbf{G}^{-1} \mathbf{B} \mathbf{x}$$

where the tilde symbol denotes mass-weighting and $\tilde{\mathbf{D}}$ is renormalized via $\mathbf{D} = \tilde{\mathbf{D}} (\mathbf{L}^T \mathbf{ML})^{1/2}$. As an important consequence, normal vibrational modes are generally delocalized over the molecule (as e.g. experienced in base pair coupling), limiting the direct use of normal mode frequencies and normal mode force constants as bond strength measure. For this purpose, local vibrational modes and related local mode frequencies and force constants are needed.

### 2.2. The local vibrational mode theory

The local vibrational mode theory, whose foundations were laid by Konkoli and Cremer [21–25], offers a solution to the delocalization problem. As shown by Konkoli and Cremer, local mode vectors $\mathbf{a}_n$ associated with the internal coordinates $q_n$, can be derived from the normal vibrational modes $\mathbf{d}_n$ and the normal mode force constant matrix $\mathbf{K}$

$$\mathbf{a}_n = \frac{\mathbf{K}^{-1} \mathbf{d}_n^T}{\mathbf{d}_n \mathbf{K}^{-1} \mathbf{d}_n^T}.$$  

where it should again be highlighted that $\mathbf{d}_n$ and $\mathbf{K}$ can be readily obtained from standard quantum chemistry packages after a vibrational frequency calculation.

The local mode force constant $k_n^a$ reads

$$k_n^a = \mathbf{a}_n^T \mathbf{K} \mathbf{a}_n = (\mathbf{d}_n \mathbf{K}^{-1} \mathbf{d}_n^T)^{-1}.$$  

from which the local mode frequency $\omega_n^a$ can be calculated

$$(\omega_n^a)^2 = \frac{4\pi^2 c^2}{m_n^a} k_n^a.$$  

The local mode mass $m_n^a$ is given by $m_n^a = (\mathbf{B}_n \mathbf{M}^{-1} \mathbf{B}_n^T)^{-1}$. 


In the context of decomposing every normal vibrational mode \( \mathbf{I}_n \) (associated with absorption peaks) into local vibrational mode percentage contributions from a non-redundant set of \( N_{vib} \) local vibrational modes \( \mathbf{a}_n \), in this respect, it is necessary to incorporate the overlapping encoded by \( S_{nu} \) [23–26] (in the discrete scalar product space),

\[
S_{nu} = \frac{\langle \mathbf{a}_n^* | \mathbf{F}^2 | \mathbf{I}_n \rangle^2}{\langle \mathbf{a}_n^* | \mathbf{F}^2 | \mathbf{a}_n^* \rangle \langle \mathbf{I}_n | \mathbf{F}^2 | \mathbf{I}_n \rangle}.
\]  

(10)

The percentage contribution \( C_{nu} \%) \) (local mode character) of local vibrational mode \( \mathbf{a}_n \) to normal vibrational mode \( \mathbf{I}_n \) reads [23]

\[
C_{nu}^\% = \frac{S_{nu}}{\sum_{m} S_{nm}} \times 100.
\]  

(11)

Recently, CNM has been successfully applied in the investigation of pK\(_a\) probes [40] and to gain insights into vibrational Stark effect probes [41].

### 2.3. Analysis of bond strength and bond nature

Zou and Cremer [42] provided the important proof that the local stretching force constant \( k^\mu(\text{AB}) \) reflects the intrinsic strength of the bond/interaction between two atoms A and B being described by an internal coordinate \( q_n \), qualifying \( k^\mu(\text{AB}) \) as perfectly suited strength descriptor. It is convenient to base the discussion for larger sets of molecules on a relative bond strength order (BSO) \( n \) rather than on a direct comparison of local force constant values. Both are connected via a power relationship of the form

\[
\text{BSO } n = A(k^\mu)^B
\]  

(12)

according to the generalized Badger rule derived by Cremer, Kraka and co-workers [43, 44]. The constants \( A \) and \( B \) are calculated from \( k^\mu \) values of two reference compounds with known BSO \( n \) values and the requirement that for a force constant value of zero the corresponding BSO \( n \) value is zero. For the HBs of the AT and GC base pairs investigated in this work, the \( A \) and \( B \) constants of equation (12) were based upon the references F-H (BSO = 1.0; \( k^\mu = 9.432 \text{ mDy} \text{n} \text{Å}^{-1} \)) and [HF-H]\(^-\) (BSO = 0.5; \( k^\mu = 0.626 \text{ mDy} \text{n} \text{Å}^{-1} \)) at the \( \omega \text{B97X-D(PCM)/aug-cc-pVTZ} \) level of theory leading to \( A \) and \( B \) constants valued at 0.564 and 0.256, respectively.

In summary, LMA has advanced over the past years as a powerful bond strength descriptor accounting for both covalent bonds and non-covalent interactions stretching from hydrogen bonds, halogen bonds to tetrel bonds and \( \pi \)-whole interactions. Some recent examples include references [45–47]. Different molecular environments were considered such as systems in solution [48] or in proteins [49]. Another focus has been metal-ligand bonding [50, 51]. The LMA of periodic systems and crystals was recently added to the repertoire [52, 53]. For a comprehensive discussion of LMA theory and its applications the reader is referred to [26] and references therein.

LMA can be complemented with an electron density analysis via Bader’s quantum theory of atoms in molecules [54–56]. In particular the covalent character of a chemical bond and/or weak chemical interaction can be assessed via the Cremer–Kraka criterion [57, 58], which is based on the local energy density \( H(r) \):

\[
H(r) = G(r) + V(r)
\]  

(13)

where \( G(r) \) is kinetic energy density (positive, destabilizing) and \( V(r) \) is potential energy density (negative, stabilizing).

If the energy density \( H(r_c) \) taken at the bond critical point \( r_c \) of the electron density \( \rho(r) \) between the two atoms A and B in question is smaller than zero, the interaction is of covalent character, if it is larger than zero, the interaction is of predominantly electrostatic character [57, 58].

### 3. Computational details

We calculated equilibrium geometries and subsequent normal vibrational frequencies of the methylated nucleobases 9-methyladenine (i.e. A), 9-methylguanine (i.e. G), 1-methylthymine (i.e. T), and 1-methylcytosine (i.e. C) and their corresponding Watson–Crick base pairs (i.e. 9-methyladenine:1-methylthymine (i.e. AT), 9-methylguanine:1-methylcytosine (i.e. GC)) (see figure I) [6], using the \( \omega \text{B97X-D} \) functional [59] in combination with Dunning's aug-cc-pVTZ basis set [60, 61]. It is
noted that all labile amino hydrogen atoms of the nucleobases and base pairs investigated were substituted with deuterium atoms [7] (see figure 1). The polarizable continuum model (PCM) [62] was utilized to create the cavity within the water dielectric medium to simulate hydration effects. Therefore, our model chemistry employs implicit solvation only. Normal vibrational mode frequencies were multiplied by an empirical scaling factor of 0.9614 to correct for anharmonicity effects [63].

Figure 2 which correlates computed normal vibrational mode frequencies (ω\text{calc}) for the nucleobases of DNA compared with 2D IR vibrational resonances (ω\text{exp}) from measurements of nucleotide monophosphates dissolved in D\textsubscript{2}O [10] proves the credibility of choice of model chemistry focusing on probing the sensitive normal vibrational modes of both the isolated nucleobases and their corresponding Watson–Crick base pairs in the fingerprint region (1500 – 1800 cm\textsuperscript{-1}).

All DFT calculations were carried out with the Gaussian 16 quantum chemistry program [64] using ultra-fine grid integration and tight convergence criterion for the self-consistent field step. LMA was performed with the standalone LModeA package [65]. Second perturbation stabilization energies [\Delta E\textsuperscript{(2)}], related to orbital interactions, were calculated using the natural bond orbital (NBO) analysis [66, 67] as implemented in the NBO program [68]. The electron density analysis was performed with the AIMAll program [69].

4. Results and discussion

Table 1 collects bond distances (d), local mode frequencies (ω\text{a}), local mode force constants (k\text{a}), electron densities [\rho(r\textsubscript{b})], energy densities [H(r\textsubscript{b})], and the ratios [H(r\textsubscript{b})/\rho(r\textsubscript{b})] calculated for the nucleobases and the corresponding Watson–Crick base pairs investigated in this work. In the following sections, the tabulated data is scrutinized.

4.1. Assessment of electronic changes in selected covalent bonds upon base pairing

In figure 3, in the context of the regimes prior to and upon base pairing, differences in distance (Δd) are compared against differences in local mode force constant (Δk\text{a}) for the C=C and C=O covalent bonds which have been suggested to partake in vibrational coupling [10, 18, 19] and the intermolecular hydrogen bonds, in both AT and GC.

Figure 3 reveals that there is a linear relationship between the changes in the distance Δd and changes in the local force constant Δk\text{a} upon base pair formation for both AT and GC base pairs (R\textsuperscript{2} = 0.9585 and
Table 1. Selected bond distances (\(d\)), local mode frequencies (\(\omega^a\)), local mode force constants (\(k^a\)), electron densities [\(\rho(r_1)\)], energy densities [\(H(r_1)\)], and the ratios [\(H(r_1)/\rho(r_1)\)] for the deuterated nucleobases and base pairs of DNA, calculated at the \(\omega\)987-X-D(PCM)/aug-cc-pVTZ level of theory in the cavity within the water dielectric medium (see text for details).

<table>
<thead>
<tr>
<th>system</th>
<th>parameter</th>
<th>(d(\text{Å}))</th>
<th>(k^a) (mDyn Å(^{-1}))</th>
<th>(\omega^a) (cm(^{-1}))</th>
<th>(\rho(r_1)) (e Å(^{-3}))</th>
<th>(H(r_1)) (E(_b) Å(^{-3}))</th>
<th>(H(r_1)/\rho(r_1)) (E(_b)/e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C13C6</td>
<td>1.404</td>
<td>6.164</td>
<td>1320</td>
<td>2.145</td>
<td>-2.429</td>
<td>-1.133</td>
</tr>
<tr>
<td></td>
<td>D18N16</td>
<td>1.003</td>
<td>7.449</td>
<td>2680</td>
<td>2.372</td>
<td>-3.820</td>
<td>-1.610</td>
</tr>
<tr>
<td></td>
<td>N14C13</td>
<td>1.340</td>
<td>6.628</td>
<td>1319</td>
<td>2.372</td>
<td>-3.312</td>
<td>-1.396</td>
</tr>
<tr>
<td></td>
<td>N14C12</td>
<td>1.336</td>
<td>6.624</td>
<td>1319</td>
<td>2.332</td>
<td>-3.397</td>
<td>-1.456</td>
</tr>
<tr>
<td></td>
<td>H15C12</td>
<td>1.084</td>
<td>5.541</td>
<td>3180</td>
<td>2.039</td>
<td>-2.332</td>
<td>-1.143</td>
</tr>
<tr>
<td>G</td>
<td>C2O1</td>
<td>1.223</td>
<td>10.870</td>
<td>1640</td>
<td>2.783</td>
<td>-4.847</td>
<td>-1.741</td>
</tr>
<tr>
<td></td>
<td>D6N3</td>
<td>1.008</td>
<td>7.275</td>
<td>2648</td>
<td>2.359</td>
<td>-3.791</td>
<td>-1.607</td>
</tr>
<tr>
<td></td>
<td>D14N9</td>
<td>1.004</td>
<td>7.345</td>
<td>2661</td>
<td>2.369</td>
<td>-3.777</td>
<td>-1.595</td>
</tr>
<tr>
<td>C</td>
<td>C2O1</td>
<td>1.229</td>
<td>10.467</td>
<td>1610</td>
<td>2.761</td>
<td>-4.790</td>
<td>-1.735</td>
</tr>
<tr>
<td></td>
<td>C9C6</td>
<td>1.348</td>
<td>8.212</td>
<td>1524</td>
<td>2.345</td>
<td>-2.993</td>
<td>-1.276</td>
</tr>
<tr>
<td></td>
<td>D15N8</td>
<td>1.003</td>
<td>7.547</td>
<td>2697</td>
<td>2.374</td>
<td>-3.823</td>
<td>-1.611</td>
</tr>
<tr>
<td></td>
<td>C5N3</td>
<td>1.322</td>
<td>7.375</td>
<td>1392</td>
<td>2.448</td>
<td>-3.554</td>
<td>-1.451</td>
</tr>
<tr>
<td></td>
<td>N3C2</td>
<td>1.356</td>
<td>6.002</td>
<td>1256</td>
<td>2.296</td>
<td>-3.096</td>
<td>-1.349</td>
</tr>
<tr>
<td>T</td>
<td>C2O1</td>
<td>1.218</td>
<td>11.266</td>
<td>1670</td>
<td>2.829</td>
<td>-4.955</td>
<td>-1.751</td>
</tr>
<tr>
<td></td>
<td>O9C5</td>
<td>1.221</td>
<td>11.154</td>
<td>1662</td>
<td>2.790</td>
<td>-4.869</td>
<td>-1.745</td>
</tr>
<tr>
<td></td>
<td>C9C7</td>
<td>1.344</td>
<td>8.496</td>
<td>1550</td>
<td>2.361</td>
<td>-3.028</td>
<td>-1.283</td>
</tr>
<tr>
<td></td>
<td>D6N3</td>
<td>1.009</td>
<td>7.286</td>
<td>2650</td>
<td>2.357</td>
<td>-3.818</td>
<td>-1.620</td>
</tr>
<tr>
<td>AT</td>
<td>C13C6</td>
<td>1.405</td>
<td>6.197</td>
<td>1324</td>
<td>2.142</td>
<td>-2.421</td>
<td>-1.130</td>
</tr>
<tr>
<td></td>
<td>D20N16</td>
<td>1.017</td>
<td>6.378</td>
<td>2479</td>
<td>2.273</td>
<td>-3.800</td>
<td>-1.672</td>
</tr>
<tr>
<td></td>
<td>N14C13</td>
<td>1.344</td>
<td>6.627</td>
<td>1319</td>
<td>2.342</td>
<td>-3.272</td>
<td>-1.397</td>
</tr>
<tr>
<td></td>
<td>N14C12</td>
<td>1.338</td>
<td>6.685</td>
<td>1325</td>
<td>2.320</td>
<td>-3.392</td>
<td>-1.462</td>
</tr>
<tr>
<td></td>
<td>H15C12</td>
<td>1.084</td>
<td>5.540</td>
<td>3180</td>
<td>2.042</td>
<td>-2.332</td>
<td>-1.142</td>
</tr>
<tr>
<td></td>
<td>C24O23</td>
<td>1.231</td>
<td>10.514</td>
<td>1613</td>
<td>2.728</td>
<td>-4.718</td>
<td>-1.730</td>
</tr>
<tr>
<td></td>
<td>C22O18</td>
<td>1.220</td>
<td>11.077</td>
<td>1656</td>
<td>2.816</td>
<td>-4.923</td>
<td>-1.748</td>
</tr>
<tr>
<td></td>
<td>C27C26</td>
<td>1.345</td>
<td>8.439</td>
<td>1545</td>
<td>2.357</td>
<td>-3.019</td>
<td>-1.281</td>
</tr>
<tr>
<td></td>
<td>N21D17</td>
<td>1.037</td>
<td>4.945</td>
<td>2183</td>
<td>2.164</td>
<td>-3.629</td>
<td>-1.677</td>
</tr>
<tr>
<td></td>
<td>O23D20</td>
<td>1.899</td>
<td>0.176</td>
<td>408</td>
<td>0.195</td>
<td>-0.007</td>
<td>-0.034</td>
</tr>
<tr>
<td></td>
<td>D17N14</td>
<td>1.840</td>
<td>0.259</td>
<td>499</td>
<td>0.275</td>
<td>-0.052</td>
<td>-0.191</td>
</tr>
<tr>
<td></td>
<td>O18H15</td>
<td>2.827</td>
<td>0.052</td>
<td>306</td>
<td>0.032</td>
<td>0.005</td>
<td>0.151</td>
</tr>
<tr>
<td>GC</td>
<td>C2O1</td>
<td>1.235</td>
<td>10.174</td>
<td>1587</td>
<td>2.708</td>
<td>-4.662</td>
<td>-1.722</td>
</tr>
<tr>
<td></td>
<td>D8N4</td>
<td>1.033</td>
<td>5.407</td>
<td>2283</td>
<td>2.196</td>
<td>-3.648</td>
<td>-1.661</td>
</tr>
<tr>
<td></td>
<td>D20N14</td>
<td>1.021</td>
<td>6.115</td>
<td>2428</td>
<td>2.240</td>
<td>-3.781</td>
<td>-1.688</td>
</tr>
<tr>
<td></td>
<td>O26C21</td>
<td>1.236</td>
<td>10.117</td>
<td>1583</td>
<td>2.719</td>
<td>-4.687</td>
<td>-1.724</td>
</tr>
<tr>
<td></td>
<td>C24C18</td>
<td>1.347</td>
<td>8.302</td>
<td>1532</td>
<td>2.351</td>
<td>-3.009</td>
<td>-1.280</td>
</tr>
<tr>
<td></td>
<td>N6D3</td>
<td>1.021</td>
<td>6.133</td>
<td>2431</td>
<td>2.252</td>
<td>-3.788</td>
<td>-1.682</td>
</tr>
<tr>
<td></td>
<td>N15C11</td>
<td>1.331</td>
<td>7.139</td>
<td>1369</td>
<td>2.395</td>
<td>-3.441</td>
<td>-1.437</td>
</tr>
<tr>
<td></td>
<td>D3O1</td>
<td>1.841</td>
<td>0.247</td>
<td>484</td>
<td>0.224</td>
<td>-0.018</td>
<td>-0.078</td>
</tr>
<tr>
<td></td>
<td>N15D8</td>
<td>1.897</td>
<td>0.475</td>
<td>677</td>
<td>0.242</td>
<td>-0.035</td>
<td>-0.146</td>
</tr>
<tr>
<td></td>
<td>O26D20</td>
<td>1.817</td>
<td>0.310</td>
<td>543</td>
<td>0.238</td>
<td>-0.024</td>
<td>-0.101</td>
</tr>
</tbody>
</table>

0.9341, respectively). In both base pairs, the most significant change occurs for the central intermolecular NH donor bond (AT: N21D27, GC: N4D8 respectively) followed by the upper NH donor bond (AT: N16D20, GC: N6D3 respectively). Clearly this reflects that the isolated nucleobases have adjusted mostly in the part forming the middle HB in the base pair.

In order to describe the same changes in bonding via the connection of a potential energy-related property with an electron density-related property, the same differences in local mode force constant (\(\Delta k^a\)) were compared against differences in energy densities [\(\Delta H(r_1)\)], which is given by figure 4 for both AT and GC. As shown in figure 4, there is a large scattering, but the main changes in \(k^a\) discussed before are accompanied by a decrease in covalent character [\(\Delta H(r_1) > 0\)] [57, 58]. Interestingly, C22018 (AT) and D20N14 (GC) exhibit a slight increase in covalent character. Again, the C=C bonds remain unchanged upon base pairing. Based on these results we conclude that the interaction between C=C and C=O bonds obviously does not cause any substantial change in C=C bond lengths nor in potential energy- and electron density-related properties upon base pairing. In the following, we will focus on the decomposition of the
characteristic normal vibrational modes via the CNM procedure, followed by the detailed analysis of the HBs involved.

4.2. CNMs
The following discussion is devoted to the consideration of the CNM analysis shown in figures 5–7, which collect the CNM plots of A, G, C, T, AT, and GC, respectively. Our results cover all normal vibrational modes within the in-plane base vibration frequency range of the nucleobases and base pairs. The active IR normal
Figure 4. Selected differences in local mode force constant ($\Delta k_a$) versus differences in energy densities [$\Delta H(r_b)$] for (a) AT and (b) GC. Calculations at the ωB97X-D(PCM)/aug-cc-pVTZ level of theory in the cavity within the water dielectric medium. C=C and C=O bonds suggested to participate in base pair coupling [10, 18, 19] are colored. (see text for details).
Figure 5. CNM plot of the characteristic vibrational region of (a) deuterated 9-methyladenine (A) and (b) deuterated 9-methylguanine (G) at the ωB97X-D(PCM)/aug-cc-pVTZ level of theory in the cavity within the water dielectric medium. Normal vibrational mode frequencies are scaled for anharmonicity effects by a factor of 0.9614. Σσ comprises the summation of all of the σ local mode contributions below the threshold of 2%.

4.2.1. Deuterated 9-methyladenine (A)

Figure 5(a) shows that the normal vibrational mode at 1596 cm\(^{-1}\) is the most localized one in the fingerprint region of A and consists of local vibrational modes whose contribution comes mainly from the pyrimidine ring, including 36% of C13C6. The subsequent normal vibrational mode at 1581 cm\(^{-1}\) has 29% of N5C2 character and exhibits some extension to the imidazole ring (8% of C2N1). This ring deformation nature agrees with the peaks denoted as A1 and A2 by Peng et al [10], respectively. However, the suggested ND\(_2\) bending contribution to the normal vibrational mode at 1596 cm\(^{-1}\) was not observed [3, 7].
4.2.2. Deuterated 9-methylguanine (G)
Based on figure 5(b), although C2O1 accounts for 62% of the normal vibrational mode at 1668 cm$^{-1}$, C4C2 and C5N3C2 together make 26% of its composition. Yet, this normal vibrational mode is the most localized one in the fingerprint region of G. Moreover, the CNM procedure confirmed the delocalized interpretation of the normal vibrational modes at 1562, 1547, and 1537 cm$^{-1}$, which is related to G2, G3, and G4, respectively [10]. The amount of delocalization increases from 1562–1537 cm$^{-1}$.

4.2.3. Deuterated 1-methylcytosine (C)
Figure 6(a) shows that the normal vibrational mode at 1636 cm$^{-1}$ has a major local mode contribution coming from 40% of C9C6, with a trace of C2O1 contribution. H10C6N4 and G6N4 add up to 26% of it as well. This is in an agreement with C2 [10]. On the other hand, the normal vibrational mode at 1625 cm$^{-1}$ contains 67% of C2O1 character, which agrees with C1 [10]. It also contains some trace of C9C6 character. Deuteration makes the equivalent of C3 vanish, and the normal vibrational modes at 1516 and 1472 cm$^{-1}$ comprise in-plane ring vibrations similarly to C4 and C5 [10].

4.2.4. Deuterated 1-methylthymine (T)
Figure 6(b) reveals that the normal vibrational modes at 1689 and 1635 cm$^{-1}$ contain major contributions from C2O1 and O9C5, revealing their interaction. The local mode characters sort of swap their percentage values from one normal vibrational mode to the other. C10C7 accounts for 53% of the normal vibrational mode at 1658 cm$^{-1}$, where minute fractions of C2O1 and O9C5 can be seen as well.

4.2.5. Deuterated 9-methyladenine:1-methylthymine base pair (AT)
Figure 7(a) allows one to quantify the increase in C=O (C24O23 and C22O18) localization of the normal vibrational modes at 1673 and 1609 cm$^{-1}$ in T. Also, there occurs a slight increase in C=C localization of the normal vibrational mode at 1565 cm$^{-1}$. The normal vibrational modes at 1599 and 1575 cm$^{-1}$ correspond to those active ones (ring deformation) in A. Of particular interest is the normal vibrational mode at 1599 cm$^{-1}$ since our CNM results reveal the interaction between C13C6 in A with the C=O groups in T, mostly C24O23. Such an interaction is also visible in a very small fraction in the normal vibrational modes at 1609 and 1575 cm$^{-1}$. These important CNM results on the interactions between those groups in AT are in perfect agreement with the literature data [15–17, 19].

4.2.6. Deuterated 9-methylguanine:1-methylcytosine base pair (GC)
Figure 7(b) shows the split of C10O1 character (G) between the normal vibrational modes at 1641 and 1630 cm$^{-1}$, where interaction with C24C18 from C is verified. On the other hand, C21O16 composes the majority of the normal vibrational mode at 1611 cm$^{-1}$ (C), with a negligible amount of C24C18 contribution. Finally, the normal vibrational modes at 1563, 1544, and 1535 cm$^{-1}$ involve ring deformation in G while that at 1516 cm$^{-1}$ contains contribution from ring deformation in C.

In summary, these results show that the characteristic C=C and C=O bond coupling suggested in the literature [10, 18, 19] can be readily probed via the CNM procedure. The degree of delocalization regarding the fingerprint region can be quantified prior to and upon base pairing. This is an important step forward toward a deeper understanding of how the building blocks of nucleic acids rearrange to form the base pairs, knowledge which could also be useful for the architecture of new so-called unnatural base pairs [70] and modified DNA base pairs with chalcogen bonding [71]. In addition, the same methodology can be employed in the future to explore base stacking effects [6] and might offer a means to probe the optical rotatory response of base complexes [72] based on a more sophisticated model chemistry.

4.3. Hydrogen bond analysis of base pairs
In subsequent we evaluated the HB interactions of the deuterated Watson–Crick base pairs in order to assess how the HB interactions modulate the strength of specific in-plane base vibrations such as C=O and N-D bonds as well as any bonds of corresponding purine (i.e. A and G and) and pyrimidine (i.e. C and T and) rings. Figure 8 shows hydrogen BSOs ($n$) calculated from local mode force constants ($k_n$). Hydrogen bond lengths ($d$) are related with the local mode force constants ($k_n$), as shown in figure 8. Also, in figure 9, the covalent character of the HB is shown as reflected by $H(r_n)$ values.

As shown in figure 8, the sensitivity of local mode force constants ($k_n$) and related BSO $n$ reveals HB2 and HB3 (non-canonical) as the strongest and weakest HBs, respectively, investigated in this work. Interestingly, the strongest hydrogen bond of AT (HB2) is roughly as strong as the weakest hydrogen bond of GC (HB1'). We found the distance of the HBs of the Watson–Crick base pairs investigated to range from
Figure 6. CNM plot of the characteristic vibrational region of (a) deuterated 1-methylcytosine (C) and (b) deuterated 1-methylthymine (T) at the ωB97X-D(PCM)/aug-cc-pVTZ level of theory in the cavity within the water dielectric medium. Normal vibrational mode frequencies are scaled for anharmonicity effects by a factor of 0.9614. Σσ comprises the summation of all of the σ local mode contributions below the threshold of 2%.

1.817 to 2.827 Å (See figure 8). Despite some scattering (R² = 0.8062), figure 8 shows the generalized Badger rule [43, 44] applies to the hydrogen bonds, i.e. the shortest hydrogen bonds (HB2' and HB3') are also the strongest whereas the longest (HB3) is the weakest. Finally, figure 9 reveals that, according to the Cremer–Kraka criterion [57, 58] for covalent bonding, all hydrogen bonds but the weakest, longest and non-canonical one (HB3) are of covalent character. This is in line with the trend that weaker hydrogen bonds have an increase in electrostatic character, as previously published by our group [73]. It is interesting to note that the corresponding hydrogen bond acceptor (C22O18) of the only electrostatic hydrogen bond (HB3)
undergoes a negative change in $H(r_b)$ upon base pairing, as shown in figure 4. It should be mentioned that for HB3 a bond critical point was detected.

Orbital interaction established between the anti-bonding $\sigma^*$ orbital of the HB donor molecule and the lone pair (i.e. lp) orbital of the HB acceptor molecule is heavily responsible for HB stabilization [48, 74]. In general, HBs become stronger via the push-pull effects that govern charge transfer events [74]. The extent of charge transfer occurring between the HB acceptor and HB donor molecules can be described in terms of

Figure 7. CNM plot of the characteristic vibrational region of (a) deuterated 9-methyladenine:1-methylthymine base pair (AT) and (b) deuterated 9-methylguanine:1-methylcytosine base pair (GC) at the $\omega$B97X-D(PCM)/aug-cc-pVTZ level of theory in the cavity within the water dielectric medium. Normal vibrational mode frequencies are scaled for anharmonicity effects by a factor of 0.9614. $\Sigma \sigma$ comprises the summation of all of the $\sigma$ local mode contributions below the threshold of 5%.
stabilization energy ($\Delta E^{(2)}$) [75]. The $\Delta E^{(2)}$ values of HB1–HB3 of AT reveal that HB2 is stronger in contrast to HB1 and HB3 due to a greater extent of charge transfer [$\Delta E^{(2)} = 13.49$ (HB1); 21.46 (HB2); 0.39 kcal mol$^{-1}$ (HB3)]. The $\Delta E^{(2)}$ values of HB1’–HB3’ of GC reveal that HB2’ is stronger than HB1’ and HB3’ due to a greater extent of charge transfer as well [$\Delta E^{(2)} = 18.22$ (HB1’); 23.32 (HB2’); 20.20 kcal mol$^{-1}$ (HB3’)]. The stabilization energies for the HBs of AT and GC are in line with figure 8(a).
Figure 9. Energy density \( [H(\tau_i)] \) versus local mode force constant \( (k^r) \) for hydrogen bonds (HBs) of base pairs AT and GC. Calculations at the \( \omega \text{B97X-D(PCM)/aug-cc-pVTZ} \) level of theory in the cavity within the water dielectric medium.

5. Conclusion

In this work, the degree of delocalization of the fingerprint normal vibrational modes, which is intimately connected to base pair coupling could be quantified and elucidated prior to and upon base pairing. This is a promising approach of detection, herein reported for the first time. We could show that the interaction between C=C and C=O bonds in nucleobases and base pairs of DNA does not cause any substantial change in the C=C bond lengths nor in potential energy-related and electron density-related properties upon base pairing. Instead, such characteristic coupling can be readily probed through the CNM procedure, an integral part of the LMA. In addition, a detailed and quantitative description of the hydrogen bonds involved in the Watson–Crick base pairs was provided. Our findings are expected to have significance to both, the experimental and theoretical vibrational spectroscopy community, and to reach out to the bioengineering community offering an efficient design tool to assess and predict the type and strength of hydrogen bonding in artificial base pairs. The protocol demonstrated within this work can be extended to future works to investigate base stacking or even to gain a better understanding of the optical rotatory response of base complexes, which would involve modified versions of the model chemistry adopted in this work.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgment

We thank SMU Center for Research Computing for providing generous computational resources. This work was financially supported by the National Science Foundation, Grant CHE 2102461. M Q thanks SMU for the Computational Science and Engineering Fellowship, as well as Bárbara M T C Peluzo for the invaluable help with figure 8(a).

ORCID iDs

Mateus Quintano  
https://orcid.org/0000-0003-4730-0337

Elfi Kraka  
https://orcid.org/0000-0002-9658-5626