Simulated NIR spectra as sensitive markers of the structure and interactions in nucleobases
Krzysztof B. Beć1*, Justyna Grabska1, Yukihiro Ozaki2, Mirosław A. Czarnecki3 & Christian W. Huck1

Near-infrared (near-IR; NIR) spectroscopy is continuously advancing in biophysical and biochemical fields of investigation. For instance, recent progresses in NIR hyperspectral imaging of biological systems may be noted. However, interpretation of NIR bands for biological samples is difficult and creates a considerable barrier in exploring the full potential of NIR spectroscopy in bioscience. For this reason, we carried out a systematic study of NIR spectra of adenine, cytosine, guanine, and thymine in polycrystalline state. Interpretation of NIR spectra of these nucleobases was supported by anharmonic vibrational analysis using Deperturbed Vibrational Second-Order Perturbation Theory (DVPT2). A number of molecular models of nucleobases was applied to study the effect of the inter-molecular interactions on the NIR spectra. The accuracy of simulated NIR spectra appears to depend on the intra-layer interactions; in contrast, the inter-layer interactions are less influential. The best results were achieved by combining the simulated spectra of monomers and dimers. It is of particular note that in-plane deformation bands are far more populated than out-of-plane ones and the importance of ring modes is relatively small. This trend is in contrast to that observed in mid-IR region. As shown, the local, short-range chemical neighborhood of nucleobase molecules influence their NIR spectra more considerably. This suggests that NIR spectra are more sensitive probe of the nucleobase pairing than mid-IR ones. The obtained results allow, for the first time, to construct a frequency correlation table for NIR spectra of purines and pyrimidines.

Nucleobases (nucleic acid bases) are one of the key building blocks of life and hence attract high attention1–3. Vibrational spectroscopy is a frequently used tool in biochemical research concerning nucleobases and their derivatives4,5. It offers high sensitivity, selectivity, structural specificity, and non-destructive sample probing6. One can study molecules not only in the gas-phase, solution, solid-state, or matrices, but also in biological samples7. These studies provide valuable information about the intermolecular interactions7, base pairing8, and tautomerization9 of nucleobases. IR and Raman spectroscopy have been used extensively in the investigations of the structure, interactions and properties of nucleobases and related compounds10–13. In contrast, investigations of near-infrared (near-IR; NIR) spectra of nucleobases are very rare14,15. Though, the experimental NIR spectra of some nucleobases have been known for a long time16, any comprehensive exploration of these spectra have not been undertaken as yet. In particular, NIR band assignments of nucleobases have never been carried out in detail, hence an understanding of NIR spectra of nucleobases not only in solutions but also in well-defined crystalline phases is still unsatisfactory. A number of open questions remain on the relation between structural and intermolecular properties of nucleobases and their NIR spectra. This knowledge is essential because the conformational flexibility of nucleobases has an important impact on their functional properties17,18.

NIR spectroscopy offers a number of advantages as compared with the other spectral regions19–21. This method provides useful information on the molecular structure, interactions, dynamics and anharmonicity22–25. However, major focus is directed towards applications of NIR spectroscopy in bioscience26, medicine27,28, chemical analysis29, and spectral imaging30–32. NIR bands are weak; typical absorption coefficients of the first overtones are smaller by up to 102 than those corresponding to mid-IR (MIR) fundamental peaks33. In the case of higher overtones, this difference may even exceed 103. This is a notable advantage of NIR spectroscopy in

1Institute of Analytical Chemistry and Radiochemistry, Leopold-Franzens University, Innsbruck 6020, Austria. 2Department of Chemistry, School of Science and Technology, Kwansei Gakuin University, Sanda, Hyogo, 669-1337, Japan. 3Faculty of Chemistry, University of Wrocław, F. Joliot-Curie 14, 50-383, Wrocław, Poland. *email: krzysztof.bec@uibk.ac.at

Corrected: Author Correction
many applications. The eventuality of a complete absorption of the radiation by a sample is far lesser, unlike in mid-IR spectroscopy that requires sample preparation or attenuated total reflection (ATR) approach. For instance, novel NIR hyperspectral imaging approaches are capable of examining entire organisms. Real-time measurements of dynamic processes such as blood flow have recently become possible. NIR spectroscopy is suitable for non-destructive investigations of the samples with high water content as well. These factors contribute to a remarkable potential of the method in biomedical applications. However, compared with mid-IR and Raman, NIR spectroscopy suffers from a poor chemical specificity, which is one of the most important problems to be addressed in current research. This is particularly significant in biological applications, in which complex spectra of biomolecules are analyzed. Therefore, the fundamental research aimed on more comprehensive understanding of NIR spectra is an important step towards increasing the chemical specificity of NIR spectroscopy.

A complexity of NIR spectra is a considerable barrier in several practical applications of NIR spectroscopy. Despite of numerous efforts, classical methods of spectral assignments in NIR region meet considerable limitations. Hence, computational chemistry is a powerful tool offering robust, detailed and independent insight into the origin of NIR bands. A remarkable progress in computational chemistry made possible for comprehensive exploration of NIR spectra. Highly accurate simulations of small- to medium-size molecules in solutions opened the way to exploration of more complex systems. Quantum chemical calculations have been applied in studies of NIR spectra of basic molecules like alcohols, phenols, nitriles, or carboxylic acids. These studies cover also a series of important molecules including melamine, thymol, rosmarinic acid, short-, medium- and long-chain fatty acids. These investigations provided interesting information on the effects of isotopic substitution, rotational isomerism, liquid structure, conformational equilibria, and hydrogen-bonding.

Highly efficient vibrational second-order perturbation theory (VPT2) is particularly useful for computational studies of medium-size or larger molecules. Therefore, this method has commonly been applied for anharmonic vibrational analyses of nucleobases and their complexes. Deperturbed and generalized variants of VPT2 method (DVPT2/GVPT2) have been adopted for detailed studies of IR features of numerous molecules, including nucleobases. For example, uracil has been examined by the second-order operator canonical Van Vleck perturbation theory (CVPT2). On the other hand, accurate but computationally highly expensive variational approach has recently been optimized in resource usage by Thomas et al. The authors applied their hierarchical intertwined reduced-rank block power method (HI-RRBPM) to uracil case. However, the previous anharmonic calculations of nucleobases have primarily been focused on IR region, and the results have been compared with the gas-phase spectra. Anharmonic calculations are gaining an increasing importance nowadays, with various methods being used for this purpose. Exploration of anharmonicity and crystal packing effects in mid-IR and Raman spectra attracts much attention because of the importance of understanding fine shifts, splitting and intensity redistribution upon the formation of weak intermolecular interactions, as for instance demonstrated by Mineeva et al. Prediction of overtones and combination bands helps an interpretation of vibrational optical activity spectra. Recently, combined approaches have been applied in Raman optical activity (ROA) and vibrational circular dichroism (VCD) spectroscopy, and computational investigations significantly extended the potential of these studies by providing insights on intermolecular interactions of chiral molecules and liquids.

Yet, these recent works focused on mid-IR, while theoretical investigations in NIR region remain rare.

The major purpose of the present study is the theoretical reproduction of NIR spectra of polycrystalline nucleobases (adenine, guanine, cytosine and thymine). In addition, we elucidate the effect of chemical neighborhood on NIR spectra of these nucleobases. The correlation between the intermolecular interactions and hydrogen-bonding of nucleobases and NIR spectra has never been attempted before. This problem is interesting in a wider context, e.g. interactions between biomolecules and water. Furthermore, we intend to explore the potential of the present approach to study important biomolecules. We expect to gain insights into the spectra-structure correlation for a well-defined molecular structure characteristic for crystal lattices. To evaluate the impact of the model complexity on the accuracy of the simulated spectra, we consider a number of molecular models (monomers, dimers, multimers). This procedure allows for indirect investigation of the effects of chemical surrounding on NIR spectra. Obtained results are used to estimate of the vibrational contributions to NIR spectra of purines and pyrimidines. On this basis, we can discuss the differences between contributions of particular modes to NIR spectra. Finally, we intend to obtain insight into the relationships between intermolecular interactions of nucleobases and their NIR features. We expect that our results will increase the chemical specificity of NIR spectroscopy and its potential in examining various biological materials (e.g. plants, natural products), tissue examination in the context of biomedical diagnosis (e.g. cancer detection), and rapidly developing NIR hyperspectral imaging of biological samples.

Methods

Experimental. Polycrystalline (powder) samples of nucleobases were purchased from Sigma-Aldrich (purity level: > 99% - guanine, cytosine, and thymine; > 98% - guanine) and milled before spectral measurements. Diffuse Reflectance (DRIFT) NIR spectra were measured on a Bruker Vector 22/N FT-NIR spectrometer. A spectrum of each sample was measured in the 10000–4000 cm⁻¹ region, with spectral resolution of 4 cm⁻¹ resulting in 2 cm⁻¹ of interpolated data spacing. For each sample, the spectral measurement was carried out three times, and 128 scans were accumulated. The spectra were measured at 298 K and then converted to absorbance scale. No spectral pretreatment was necessary.

Computational details. Multi-modal anharmonic computations were necessary in order to obtain data on the first overtones and binary combinations, which are the most influential contributions into NIR spectra. Due to an extensive computational cost of anharmonic approximation, it was required to represent the major structural features of crystalline nucleobases by simplified molecular models. We undertook a systematic approach for each nucleobase we selected a number of structures of increasing complexity, from monomers to clusters.
consisting of up to six molecules. The initial molecular geometries of the clusters were extracted from the crystal structures available from Cambridge Structural Database (CSD)\(^7\). Further details of this approach will be shown in Results and Discussion Section.

Computations of NIR spectra were performed using Deperturbed Vibrational Second-Order Perturbation Theory (DVPT2) approach\(^4\)\(^3\)\(^,\)\(^4\)\(^4\). The determination of the ground-state electronic properties was conducted with Density Functional Theory (DFT) calculations at M06-2X level. M06-2X single-hybrid density functional is parameterized towards long-range interactions\(^7\)\(^4\), which were additionally refined by applying Grimme’s third version of empirical correction for dispersion (GD3)\(^7\)\(^5\). The following basis sets were applied: 6-311++G(2df, 2pd) for monomers, and 6-31 G(d, p) for dimers and larger clusters. Additionally, these calculations were also done with the use of B3LYP functional. The comparison between M06-2X and B3LYP approaches revealed that the latter strongly overestimates the inter-layer interactions between nucleobase molecules in the clusters. This leads to flattening of the two-layered models into single-layer structures. On the other hand, ground state structures obtained with M06-2X functional relatively well reflects the crystalline state. Therefore, only the results of M06-2X calculations are discussed in this work. All quantum mechanical calculations were performed with Gaussian 09 Rev. E01 software\(^7\)\(^6\).

The spectral lineshapes were modelled by Lorentz-Gauss (Cauchy-Gauss) product function\(^7\)\(^7\)\(^,\)\(^7\)\(^8\). The processing and visualization of the experimental and calculated data was done using script written in MATLAB (The MathWorks Co.).

**Results and Discussion**

**Structural simulation of NIR spectra of solid state nucleobases.** Chemical environment in a crystalline lattice is highly specific and different from that of non-interacting molecules diluted in solvents\(^2\)\(^3\) or well-defined complexes formed in solutions\(^8\)\(^9\). On the other hand, our previous investigation of polycrystalline melamine suggested that the simplification of molecular models does not reduce remarkably the accuracy of calculated overtones and combinations\(^5\)\(^0\). To some extent, the structure of melamine resembles that of the nucleobases. However, it is necessary to examine whether this similarity also appears in NIR and mid-IR vibrational features\(^5\)\(^0\). To verify this assumption, at first we calculated NIR spectra using single molecule models of adenine, guanine, cytosine, and thymine. This approximation completely neglects the chemical surrounding effect. Intermolecular interactions of nucleobases may roughly be separated into in-plane and inter-plane ones\(^7\)\(^9\). The former ones are stronger and predominantly controlled by hydrogen-bonding between the proton donor and acceptor centers of nucleobases\(^8\). The latter ones include ring stacking forces, which are relatively weaker\(^4\)\(^0\). We used the dimers as the simplest model that mimic in-plane (or XY plane) interactions (Fig. 1). We selected the most stable dimeric structures known in the literature\(^8\)\(^1\). To gain insight into the role of both kinds of interactions, at least four molecules in two layers are required (Fig. 2). The layered models include inter-layer (or Z plane) interactions as well.

Thus, we constructed four-molecules clusters by extracting their geometry from the crystalline structures. Prior to the vibrational analysis, it was mandatory to optimize these structures to reach the minimum at the potential energy surface. This step resulted in a modest distortion of the clusters in comparison with the crystalline structures. To minimize this effect, and keep the computational cost at an acceptable level, we have selected M06-2X-GD3/6-31 G(d, p) method that is computationally efficient and reasonably well describes the inter-molecular forces existing between the molecules of nucleobases. The obtained clusters adequately reproduced the major structural motifs existing in crystal lattices (Fig. 2). Molecules of pyrimidines are smaller and allow for calculation of considerably larger clusters. Thymine cluster consisting of six molecules was still acceptable for the anharmonic treatment. Unlike the case of cytosine, the larger cluster of thymine was only slightly distorted during the optimization. For this reason, we selected the six-molecule cluster of thymine for the present study (Fig. 2).

---

**Figure 1.** Dimer structures of (A) adenine; (B) cytosine; (C) guanine; (D) thymine.
We considered a number of factors to improve the agreement between the simulated and experimental NIR spectra of crystalline nucleobases. Even simple inspection of the experimental spectra reveals a pronounced baseline elevation (Fig. 3A–D). This effect is observed in a broad spectral region (from 6700 to 4000 cm$^{-1}$) for all four compounds. Similar effect was observed in NIR spectra of carboxylic acids. These studies also evidenced the concentration dependence of the baseline fluctuation. As in the case of carboxylic acids, one can notice an existence of at least two components contributing to the observed baseline fluctuation. This effect is the most pronounced in the spectrum of thymine (Fig. 3D) and adenine (Fig. 3A), while in the spectrum of cytosine (Fig. 3B)
and guanine (Fig. 3C) the baseline is more uniformly elevated throughout entire NIR region. It is probable that the appearance of both components is related to strong hydrogen–bonding interactions present in the samples. However, the phenomenon responsible for the background elevation was not reproduced in our calculations. The main origin of the disagreement between the calculated and experimental values of band positions and intensities at present, we are unable to propose an exhaustive explanation of this effect. It is interesting to note that such broad spectral features appear in NIR spectra of compounds that form cyclic dimers with strong hydrogen bonding. One may speculate that the delocalization of the electron density is responsible for this broad absorption. However, additional studies are necessary to confirm this hypothesis. To improve the agreement between the experimental and theoretical spectra, the baseline was numerically fitted. For this purpose, we applied the same band model as that used for modeling of NIR bands. Details of this procedure are presented in Supporting Information.

As expected, distinct differences in the theoretical NIR spectra appear when using either the monomers, dimer, or clusters for the calculations. Obviously, the models of monomers do not include intermolecular interactions, and thus are expected to be insufficient. Surprisingly, the models of clusters develop a number of additional bands, which are not present in the experimental spectra. The theoretical spectra combining monomers and dimers provided the best approximation of the experimental spectra. This conclusion is in line with the former simulations of NIR spectra of short- and medium-chain fatty acids. Here, by mixing the theoretical spectra of the dimer and monomer with 1:1 ratio we obtained a reasonably accurate reproduction of NIR spectra of crystalline nucleobases. To improve the analysis of the spectra, we applied uniform wavenumber scaling for each of the calculated spectra (adenine: 0.9811; cytosine: 0.9854; guanine: 0.9615; thymine: 0.9807). The spectra adjusted this way were used for further discussions. Note that the empirical frequency scaling factors are intended to correct all kinds of errors resulting from approximations that are necessary in quantum chemical calculations. The purpose of the scaling factors is to adjust the calculated frequencies to match the experimental ones. In this study, the application of scaling factors followed an observation that the meaningful calculated peaks have systematically overestimated positions. Hence, scaling was applied for better presentation of the spectra and to ease the discussion. A few factors may contribute to the observed overestimation of the calculated peak positions. Primarily, the inaccuracies in the potential energy resulting from approximations in the electronic theory used in our calculations, density functional theory and basis set combination. Moreover, this effect may also result from properties of VPT2 theory. Its computational efficiency is obtained through a relatively shallow probing of the vibrational potential. VPT2 approach predicts the shape of the potential based on its local curvature. This may result in an overestimation of X–H frequencies, e.g. as reported by us previously.

Contribution of various vibrational modes to NIR spectra of nucleobases. The calculated NIR spectra include the first overtones and binary combinations. Our previous studies have shown that the contributions from higher quanta transitions to NIR spectra in the 10,000–4000 cm\(^{-1}\) region (e.g. second overtones and ternary combinations) are smaller than 20%. Thus, the majority of spectral information in this range is well-reflected by the first overtones and binary combinations. As mentioned before, NIR spectra are complex with magnitude of overlapping peaks. To ease the spectral analysis of nucleobases, we used projections of the relative contributions of pre-selected modes as a function of wavenumber.

Figures 4–7 present the contributions of selected vibrational modes to simulated NIR spectra of crystalline nucleobases. The first overtones \((2\nu)\) contribute predominantly in the 7500–5500 cm\(^{-1}\) region. As expected, the largest contributions should be assigned to the NH and NH\(_2\) stretching modes. In the region of 6500–5500 cm\(^{-1}\) one can also observe the combination bands resulting from stretching vibrations.

The region of 5800–4000 cm\(^{-1}\) is contributed almost entirely by the combination bands \((\nu_x + \nu_y; \nu_x, \nu_y\) refer to any given stretching or bending mode). The most important are the combinations involving the \(\nu_{\text{NH}}\) and \(\nu_{\text{NH}_2}\) vibrations. The relative contributions are similar for all four nucleobases. As expected, the importance of the \(\nu_{\text{C–O}}\) combination bands is relatively low in NIR spectra, in contrast to mid-IR spectra of nucleobases, where these bands strongly absorb in the 1735–1667 cm\(^{-1}\) region. Interestingly, thymine has two C–O moieties in its structure and the predicted \(\nu_{\text{C–O}}\) contribution to NIR spectrum (Fig. 7) is higher than that for cytosine (Fig. 6) and guanine (Fig. 5). In addition to the \(\nu_{\text{C–O}}\) combination bands, further studies are necessary to explain a possible contribution from the second overtone of the C–O stretching vibration \((3\nu_{\text{C–O}})\). This band appears as a moderately intense feature near 5100 cm\(^{-1}\) in other molecules.

The contributions of the combinations involving \(\delta_{\text{NH}_2}\) vibrations are significant for all four nucleobases. Rocking \(\text{NH}_2\) vibration is noticeable in the spectra of adenine and cytosine, but less evident in guanine. Bands from wagging and twisting \(\text{NH}_2\) vibrations do not appear in NIR spectra of nucleobases, except for cytosine, for which the contribution from the twisting mode is observed near 4000 cm\(^{-1}\) (Fig. 6).

The contributions from the NH deformation modes depend on the symmetry of vibration. In-plane vibrations are significant for all molecules, while out-of-plane ones are important only in the spectra of thymine (Fig. 7) and guanine (Fig. 5). The contributions from the in-plane CH\(_3\) deformation modes are more significant than those from the out-of-plane vibrations. It is expected since the in-plane deformation bands appear at higher wavenumbers than the out-of-plane ones. It is interesting to note that the contributions from the in-plane vibrations are stronger for nucleobases with purine rings (adenine, guanine) as compared with the nucleobases with the pyrimidine ring (cytosine, thymine). Similar trend one can observe for in-plane and out-of-plane ring vibrations. The in-plane vibrations are more important than the out-of-plane ones. In addition, the contributions from the in-plane vibrations are more significant for purines. These observations confirm a strong relationship between the molecular structure and the NIR spectrum.

Thymine is particularly intriguing because it features a CH\(_3\) group in its structure (Figs 1, 7). As can be seen, the contribution of the CH\(_3\) vibrations to NIR spectra remains moderate. The first overtones of the CH\(_3\) stretching modes contribute in relatively narrow wavenumber regions (Fig. 7). One can also notice that the combinations
involving the $\nu CH_3$ and $\delta CH_3$ modes are rather weak (Fig. 7). Additionally, these bands are heavily overlapped with more pronounced combinations from the $\nu XH$, $\nu C=O$, and $\delta NH$ modes.

**NIR bands as vibrational markers of nucleobases.** Systematic studies of mid-IR and Raman spectra resulted in establishing of the correlation tables and characteristic bands originating from purines and pyrimidines. Therefore, nucleobases and their derivatives (e.g. corresponding nucleotides) can be discriminated in mid-IR and Raman spectra of biological samples. The bands appearing in the lower mid-IR region are particularly useful. For instance, Mello *et al.* have reported discrimination between the four nucleobase units in the spectra of DNA. The C=O stretching fundamental frequencies of guanine appear at 1710 and 1716 cm$^{-1}$, while those of thymine are observed at 1700 and 1664 cm$^{-1}$. Banyay *et al.* have reported a list of characteristic wavenumbers that provide highly specific information on nucleic acid structure, including pairing and stacking effects. Band shift and intensity changes in response to various effects, like conformational changes, intermolecular interactions or solvent effect are available in the literature. Comparable libraries of Raman characteristic frequencies are accessible as well. In contrast, similar data are not available for NIR region. Our calculations allow, for the first time, to establish the spectra-structure correlations in NIR region for nucleobases (Table 1).

---

**Figure 4.** Comparison of the experimental and theoretical (DVPT2//M062X/6–311++G(2df, 2pd) NIR spectra of adenine together with the theoretical contributions of selected vibrational modes to NIR intensity.
We identify a number of specific combination bands of purines and pyrimidines (Table 1). This provides an opportunity to discriminate both structures in NIR spectra. From the point-of-view of chemical specificity, the overtone bands appear to be less useful. This results from the red-shift and broadening of the overtone bands due to hydrogen-bonding. The resulting broad, overlapping bands may become even less distinct in the spectra of biological samples where overlapping with the $\nu\text{CH}$ bands of lipids is anticipated\textsuperscript{36–38}.

It seems that the combination bands involving the NH and NH$_2$ vibrations are the most essential for identification of nucleobases. These bands have relatively high contributions to NIR spectra and are less likely to be obscured by the bands from other biomolecules, which makes them particularly useful in applications. The NH$_2$ groups in purines (adenine, guanine) and cytosine develop sharp peaks in vicinity of 5030–4970 cm$^{-1}$. NIR bands associated with the NH stretching and deformation vibrations are particularly specific for thymine. Besides, only thymine shows characteristic stretching bands due to the CH$_3$ group (Table 1). These vibrations give rise to a distinct, sharp and intense doublet at 4426 and 4332 cm$^{-1}$ (Fig. 7).

**Figure 5.** Comparison of the experimental and theoretical (DVPT2//M062X/6-311++G(2df, 2pd)) NIR spectra of guanine together with the theoretical contributions of selected vibrational modes to NIR intensity.
General impact and possible applications of the present study. The present study demonstrates the feasibility of obtaining detailed insight into the origin of NIR bands of nucleobases through the quantum mechanical calculations. This approach helps to distinguish the bands of nucleic acids in NIR spectra of biological samples. Improvement of the chemical specificity and the capability of yielding structural insights is expected to enhance the potential of NIR spectroscopy in bioscience. For the first time, it was possible to establish a correlation table linking the frequencies of the characteristic NIR bands of nucleobases with their structural fragments.

Moreover, our results evidence that NIR bands are more sensitive to the effect of the nearest chemical environment than to the longer-range structural arrangement. This is in contrast to the trends observed for fundamental bands in mid-IR and Raman spectra. Calculations based on relatively simple molecular models of nucleobases tend to reproduce accurately the experimental NIR bands in crystalline state. This observation is in contrast to simulated mid-IR spectra, where the molecular complexity needs to be reflected by the model much more precisely. This creates an opportunity for efficient theoretical simulation of NIR spectra of other kinds of biomolecules.

Figure 6. Comparison of the experimental and theoretical (DVPT2//M062X/6-311++G(2df,2pd)) NIR spectra of cytosine together with the theoretical contributions of selected vibrational modes to NIR intensity.
The observed relatively lower sensitivity of NIR bands to the long-range interactions has a far-reaching impact. NIR spectra of nucleobases are sensitive to the local neighborhood of the molecule, but less sensitive to a more remote chemical environment. This observation suggests that NIR bands are better probes of e.g. nucleobase pairing, by being relatively less obscured by other effects when compared with mid-IR and Raman spectra. The hydrogen-bonding between nucleobases results in characteristic NIR spectral pattern including red-shift and substantial broadening of the overtone bands of NH and NH$_2$ groups. This effects likely contribute to the profound baseline elevation in NIR spectra of nucleobases. Similar effect has also been observed in the spectra of fatty acids$^{53-55}$. In contrast, combination bands, even those that involve $\nu$NH/NH$_2$ vibrations, are relatively less affected than the overtones.

**Conclusions**

NIR spectra of adenine, guanine, cytosine, and thymine were reproduced by anharmonic calculations. The obtained results allow for detailed insight into the origin of NIR bands, the relationship with molecular structure and intermolecular interactions. The simulated first overtones and binary combination bands have been compared with the NIR spectra measured for polycrystalline samples. The calculations were based on a number of molecular models including monomers, dimers, and clusters consisting of four to six molecules. The most
Contributing to NIR spectra. We have demonstrated that neither combination bands nor overtones of defor-
X-H stretching vibrations. On the other hand, a very limited information is available on the other kinds of bands
troscopy may be used as an efficient tool for monitoring which substances cause direct mutations in living cells.
formation modes follow the pattern known for the overtones of stretching vibrations. Additionally, the profound
baseline elevation effect was noticed. Similar effect was previously observed in NIR spectra of fatty acids. Two
mation modes with the other modes influence considerably the region of 5200–4000 cm−1. The overtones and
combinations involving νCH, δCH Adenine 6148–6103
combinations involving νCH, δCH Cytosine 6206
combinations involving νCH, δCH Thymine 4206
combinations involving νCH, δCH, ννCH, νδ ring Guanine 4908
combinations involving νCH, δCH, νδ ring Cytosine 4678
combinations involving νCH, δCH, νδ ring Thymine 4650
combinations involving νCH, δCH, ννCH, νδ ring Guanine 4972–4946
combinations involving νCH, δCH, νδ ring, δν NH Cytosine 4977
combinations involving νδ NH, δν δring Guanine 4998
combinations involving νδ NH, δν δring, δν NH Cytosine 5997
combinations involving νδ NH, δν δring Thymine 5997
combinations involving νδ NH, δν δring Cytosine 5992
combinations involving νδ NH, δν δring, νν NH Thymine 6149–6082
combinations involving νδ NH, δν δring, νν NH, νδ CH Cytosine 6148–6103
combinations involving νδ NH, δν δring, νν NH, νδ CH Thymine 6670–6450
combinations involving νδ NH, δν δring, νν NH, νδ CH Guanine 6978–6720
combinations involving νδ NH, δν δring, νν NH, νδ CH Cytosine 6940–6025
combinations involving νδ NH, δν δring, νν NH, νδ CH Thymine 5890–5890

Table 1. Characteristic NIR bands of nucleobases.

Table 1. Characteristic NIR bands of nucleobases.

accurate approximation of the experimental spectra was accomplished by a combination of the theoretical spectra of
the monomers and dimers with a subsequent wavenumber scaling.
Our results provide new insight into the origin of NIR bands of nucleobases. It appears that for all studied
nucleobases the combinations bands are far more important than the overtone bands. The overtones of the
NH stretching vibrations strongly contribute to the higher wavelength region of NIR spectra, while their binary
combinations with the other modes influence considerably the region of 5200–4000 cm−1. The overtones and
combination bands of low-lying fundamentals do not appear in NIR spectra of nucleobases, e.g. out-of-plane ring
deformation modes. This is in agreement with the observation that the cluster models that include inter-plane
interactions do not improve the accuracy of the simulated spectrum. Thus, the in-plane NIR modes of nucleo-
bases are well approximated by model including the dimers only.
It is of note that more complex models of nucleobases provide poorer results as compared to smaller models.
This result corresponds well with that obtained for melamine50. The exact description of crystalline state appears
to be less important for obtaining of an accurate theoretical NIR spectrum than for the corresponding mid-IR
spectrum9. The present study confirms that NIR bands are relatively less sensitive for the intermolecular inter-
action and chemical environment compared with the fundamental bands. This observation suggests that NIR spec-
troscopy may be used as an efficient tool for monitoring which substances cause direct mutations in living cells.
As yet, most of NIR spectroscopic studies of intermolecular interactions have been focused on overtones of
X-H stretching vibrations. On the other hand, a very limited information is available on the other kinds of bands
contributing to NIR spectra. We have demonstrated that neither combination bands nor overtones of deformation
modes follow the pattern known for the overtones of stretching vibrations. Additionally, the profound
baseline elevation effect was noticed. Similar effect was previously observed in NIR spectra of fatty acids. Two
broad spectral features seem to be responsible for this effect. At present, the origin of these two components is not
known and additional studies are required.
The correlation tables between NIR spectra and the structure of nucleobases obtained from the theoretical
simulations is an important step towards expanding of potential of NIR spectroscopy in biomedical applications.

References
1. Miescher, F. Ueber die chemische Zusammensetzung der Eiterzellen. Medicinisch-chemische Untersuchungen. 4, 441–460 (1871).
2. Heard, R. D., Kinnersley, H. W., O’brien, J. R., Peters, R. A. & Reader, V. Vitamin B1 and adenine. Nature 131, 617–618 (1933).
3. Xu, W., Chan, K. M. & Kool, E. T. Fluorescent nucleobases as tools for studying DNA and RNA. Nat. Chem. 9, 1043–1055 (2017).
4. Barth, A., Haris, P. I. Biological and biomedical infrared spectroscopy, Vol. 2 Advances in Biomedical Spectroscopy. IOS Press BV, Amsterdam, Netherlands (2009).
5. Taillandier, E. & Liquier, J. Infrared spectroscopy of DNA. Methods Enzymol. 211, 307–335 (1992).
6. Chalmers, J. M. & Griffiths, P. R. (Ed.). Handbook of vibrational spectroscopy. John Wiley & Sons, Chichester (2002).
7. Stuart, B. H. Infrared spectroscopy of biological applications: an overview. John Wiley & Sons, New York (2011).

The correlation tables between NIR spectra and the structure of nucleobases obtained from the theoretical
simulations is an important step towards expanding of potential of NIR spectroscopy in biomedical applications.

Received: 25 August 2019; Accepted: 11 October 2019;
Published online: 22 November 2019

Table 1. Characteristic NIR bands of nucleobases.

| Characteristic bands (major contributions) | Nucleobase | Wavenumber (cm⁻¹) |
|------------------------------------------|------------|------------------|
| combinations involving νNH               | Thymine    | 4080             |
| combinations involving νCH               | Cytosine   | 4206             |
| combinations involving νNH, δν NH, νν CH | Thymine    | 4354–4332        |
| combinations involving νCH, δν NH, δρ ring| Adenine    | 4361             |
| combinations involving νν NH, δν NH, δρ ring| Cytosine | 4426             |
| combinations involving δν NH, δρ ring    | Guanine    | 4493             |
| combinations involving νδ NH, νδ NH, δρ ring| Cytosine | 4519–4481        |
| combinations involving νδ NH, νδ NH, δρ ring| Adenine | 4528             |
| combinations involving νδ NH, νδ NH, δν CH | Thymine    | 4550             |
| combinations involving δν CH, δρ ring    | Adenine    | 4768             |
| combinations involving νδ NH, νδ NH, δν CH | Guanine    | 4908             |
| combinations involving δν NH, δν CH, δρ ring| Adenine | 4972–4946        |
| combinations involving δν NH, δν CH, δρ ring, δν NH | Cytosine | 4997             |
| combinations involving δν NH, δν CH, δρ ring, δν NH | Guanine | 4998             |
| combinations involving δν CH, νν CH | Thymine    | 5997             |
| combinations involving νNH               | Cytosine   | 5992             |
| combinations involving νCH               | Adenine    | 6149–6082        |
| combinations involving νNH, νν CH | Guanine    | 6148–6103        |
| overtones of νNH                         | Adenine    | 6670–6450        |
| overtones of νNH                         | Cytosine   | 6978–6720        |
| overtones of νNH, νν CH                  | Guanine    | 6940–6025        |
| overtones of νNH, νν CH                  | Thymine    | 5890–5890        |
52. Kirchler, C. G.
54. Grabska, J., Beć, K. B., Ishigaki, M., Woźnica, M. J. & Ozaki, Y. Spectroscopic and quantum mechanical calculation study of the effect of isotopic substitution on NIR spectra of methanol. J. Phys. Chem. A 121, 7925–7936 (2017).
56. Grabska, J., Czarnecki, M. A., Beć, K. B. & Ozaki, Y. Structure and near-infrared spectra of saturated and unsaturated carboxylic acids. An insight from anharmonic DFT calculations. J. Phys. Chem. A 121, 3437–3451 (2017).
58. Bloino, J., Baiardi, A. & Biczysko, M. Aiming at an accurate prediction of vibrational and electronic spectra for medium-to-large molecules: An overview. Int. J. Quantum Chem. 116, 1543–1574 (2016).
60. Puzzarini, C., Biczysko, M. & Barone, V. Accurate anharmonic vibrational frequencies for uracil: the performance of composite schemes and hybrid CC/DFT model. J. Chem. Theory Comput. 7, 3702–3710 (2011).
64. Krasnoshchekov, S. V., Vogt, N. & Stepanov, N. F. Ab Initio anharmonic analysis of vibrational spectra of uracil using the numerical-analytic implementation of operator Van Vleck perturbation theory. J. Phys. Chem. A 119, 6723–6737 (2015).
66. Minaeva, V. A., Minaev, B. F., Baryshnikov, G. V., Romeyko, O. M. & Pittelkow, M. The FTIR spectra of substituted tetrathia[8]circulenes and their assignments based on DFT calculations. Vibrational Spectroscopy 65, 147–158 (2013).
68. Minaeva, V. A., Minaev, B. F., Baryshnikov, G. V., Romeyko, O. M. & Pittelkow, M. The FTIR spectra of substituted tetrathia[8]circulenes and their assignments based on DFT calculations. Vibrational Spectroscopy 65, 147–158 (2013).
70. Pecul, M. & Sadlej, J. Chiral Recognition by Molecular Spectroscopy. In Molecular spectroscopy: a quantum chemistry approach. ed. J. Chem. Phys. A 121, 1543–1574 (2016).
72. Grimm, S., Antony, I. Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. J. Chem. Phys. 132, 154104 (2010).
74. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. Theor. Chem. Acc. 120, 215–241 (2008).
76. Gaussian 09, Revision E.01, Frisch, M. J. et al. Gaussian, Inc., Wallingford, CT (2013).
78. Grabska, J., Ishigaki, M., Beć, K. B., Woźnica, M. J. & Ozaki, Y. Correlations between structure and near-infrared spectra of saturated and unsaturated carboxylic acids. Insight from anharmonic density functional theory calculations. J. Phys. Chem. A 121, 3437–3451 (2017).
80. Sigel, A. & Krylov, A. I. Scaling factors and uncertainties for ab initio anharmonic vibrational frequencies. J. Chem. Phys. 100, 1511–1518 (1994).
84. Mathlouthi, M. & Seuvre, A. M. FT-I.R. and laser-Raman spectra of thymine and thymidine. Carbohydrate Research 134, 23–38 (1984).
86. Mathlouthi, M. & Seuvre, A. M. FT-I.R. and laser-Raman spectra of guanine and guanosine. Carbohydrate Research 146, 15–27 (1986).
87. Mello, M. L. S. & Vidal, B. C. Changes in the infrared microspectroscopic characteristics of DNA caused by cationic elements, different base richness and single-stranded form. *PLoS One* **7**, e43169 (2012).
88. Banyay, M., Sarkarb, M. & Graslund., A. A library of IR bands of nucleic acids in solution. *Biophysical Chemistry* **104**, 477–488 (2003).
89. Movasaghi, Z. & Rehman, S. ur Rehman, I. Fourier transform infrared (FTIR) spectroscopy of biological tissues. *Appl. Spectrosc. Rev.* **43**, 134–179 (2008).
90. Olszyńska-Janus, S., Gąsior-Głogowska, M., Szymborska-Malek, K., Czarnik-Matusewicz, B. & Komorowska, M. Specific applications of vibrational spectroscopy in biomedical engineering. *IntechOpen*, https://doi.org/10.5772/13121 (2011).

**Acknowledgements**

This work was funded by the Austrian Science Fund (FWF): M2729-N28. This work was partially supported by the National Science Center Poland, Grant 2016/21/B/ST4/02102. Calculations have been carried out in Wroclaw Centre for Networking and Supercomputing (http://www.wcss.pl), under Grant no. 163.

**Author contributions**

K.B.B. conceived and designed this research. K.B.B. and J.G. performed experimental and computational study. K.B.B. and J.G. analyzed the data and draw conclusions with the help from Y.O., M.A.C. and C.W.H. K.B.B. prepared the manuscript. All authors took part in scientific discussion of the results, manuscript review, editing and improvements.

**Competing interests**

The authors declare no competing interests.

**Additional information**

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-53827-6.

Correspondence and requests for materials should be addressed to K.B.B.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2019