Modeling the amide I bands of small peptides
Jansen, Thomas la Cour; Dijkstra, Arend G.; Watson, Tim M.; Hirst, Jonathan D.; Knoester, Jasper
Published in: Journal of Chemical Physics
DOI: 10.1063/1.2218516

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2006

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Modeling the amide I bands of small peptides

Thomas la Cour Jansen and Arend G. Dijkstra
Institute for Theoretical Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands and Materials Science Centre, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Tim M. Watson and Jonathan D. Hirst
School of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom

Jasper Knoester
Institute for Theoretical Physics, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands and Materials Science Centre, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

(Received 20 March 2006; accepted 6 June 2006; published online 28 July 2006)

In this paper different floating oscillator models for describing the amide I band of peptides and proteins are compared with density functional theory (DFT) calculations. Models for the variation of the frequency shifts of the oscillators and the nearest-neighbor coupling between them with respect to conformation are constructed from DFT normal mode calculations on N-acetyl-glycine-N’-methylamide. The calculated frequencies are compared with those obtained from existing electrostatic models. Furthermore, a new transition charge coupling model is presented. We suggest a model which combines the nearest-neighbor maps with long-range interactions accounted for using the new transition charge model and an existing electrostatic map for long-range interaction frequency shifts. This model and others, which account for the frequency shifts by electrostatic maps exclusively, are tested by comparing the predicted IR spectra with those from DFT calculations on the pentapeptide [Leu]-enkephalin. The new model described above gives the best agreement and, after a systematic blueshift is accounted for, reproduces the DFT frequencies to within 3.5 cm\(^{-1}\). The correlation of the intensities for this model with intensities from DFT calculations is 0.94. © 2006 American Institute of Physics. [DOI: 10.1063/1.2218516]

I. INTRODUCTION

The amide I region of the infrared spectra of proteins and peptides has received much attention.\(^1\)–\(^19\) This region is particularly interesting, because the absorption is strong. The amide I vibration is dominated by the CO stretch found in every amino acid. The amide I vibrations of different residues are strongly coupled and the coupling depends on the structure of the protein. For \(\alpha\) helices this leads to a single peak at 1650 cm\(^{-1}\), while two peaks are observed for \(\beta\) sheets at 1640 and 1680 cm\(^{-1}\).\(^12\) This makes it possible to use infrared spectroscopy to observe structural changes in solution. More detailed information can be extracted using isotope labeling\(^15\),\(^17\) and/or two-dimensional infrared correlation spectroscopy.\(^7\)–\(^9\),\(^16\),\(^19\)

The water bending vibration gives an absorption in the same frequency range as the amide I mode. Infrared measurements are therefore commonly performed in heavy water, moving the water bending frequency to the 1200 cm\(^{-1}\) region. Solvation in heavy water leads to the exchange of acidic hydrogen with deuterium. Among the hydrogen atoms exchanged are those on the nitrogen atom of the backbone amide groups.

For small molecules in the gas phase, vibrational spectra can be calculated within an accuracy of a few percent by finding the normal modes using \(ab\ initio\) or density functional theory (DFT) calculations.\(^20\) The error is dominated by a systematic shift, due to the neglect of correlation with ionic bond structures at bond elongation. The frequency differences computed for similar structures will therefore be accurate to within a few wave numbers. However, for whole proteins in solution, another approach is needed for theoretical investigations on the amide I band. The floating oscillator model\(^2\) assumes that the amide I vibrations do not mix with other vibrations and that they can be described by an exciton Hamiltonian involving an amide I oscillator localized on each amide bond and the coupling between these floating oscillators.

This leads to the construction of the exciton Hamiltonian

\[
H = \sum_{i=1}^{N} \left( \omega_i B_i^\dagger B_i + \sum_{j=1}^{N} J_{ij} B_i^\dagger B_j \right)
\]  

(1)

\(B_i\) and \(B_i^\dagger\) are Bosonic annihilation and creation operators fulfilling the commutation relation \([B_i, B_j^\dagger] = \delta_{ij}\). \(\omega_i\) is the fundamental frequency of site \(i\), while \(J_{ij}\) is the coupling between site \(i\) and site \(j\). In order to obtain the infrared spectrum one needs these parameters as well as the transition dipole \(\mu_i\) for each site.

The goal of this paper is to construct a new set of models for these parameters for the Hamiltonian and validate these and other models by comparing the calculated infrared spec-
for the present study. We have selected the following set of maps under estimated by about 25 cm$^{-1}$. We will compare the site frequencies that are obtained using the partial charges from the GROMOS96 (Ref. 30) and OPLS (Ref. 31) force fields combined with the three maps described above.

The second set of parameters needed for the floating oscillator models are the couplings between amide I vibrations of different sites. This coupling was first modeled using the transition dipole coupling (TDC) model. Torii and Tasumi showed that this model is inadequate for describing the coupling of neighboring peptide units. They constructed a map of the coupling as function of the Ramachandran angles between the neighboring peptide units (the Tasumi map). This was done using RHF calculations on N-acetyl-glycine-N'-methylamide, also known as the glycine dipeptide (GLDP). This type of map is reasonably transferable. Hamm and Woutersen suggested a transition charge coupling (TCC) model, which improves on the TDC model by including higher-order multipole contributions. The model agreed reasonably well with the coupling constant calculated with DFT on GLDP. The remaining discrepancy between the DFT calculation and the TCC model can mainly be attributed to through-bond coupling, which cannot be described with an electrostatic model, such as TCC. In the Appendix we review the electrostatic coupling models and in Sec. II a new TCC model based on multipole derived charges (MDC) will be proposed.

In this paper a model for the nearest-neighbor frequency shifts and couplings will be constructed. This is done in an analogous way to the maps constructed for the couplings as described above. The frequencies of the two amide sites of GLDP, denoted the N and C sites (see Fig. 1), are extracted along with the coupling between the two units from DFT normal mode calculations on GLDP with different Ramachandran angles. The site frequencies and couplings are compared with those obtained with electrostatic maps and existing coupling models. The nearest-neighbor coupling model will be combined with electrostatic maps and TDC and TCC models to account for the long-range interactions.

In order to test the new models, we chose the pentapeptide [Leu]-enkephalin, a neurotransmitter involved in pain regulation. (amino acid sequence: YGGFL, Fig. 1). [Leu]-enkephalin has four amide sites numbered from 1 to 4 starting from the N-terminus. The gas phase IR spectra of the ten peptides. This will allow us to estimate the accuracy one can expect for larger systems. Furthermore, this will help us identify challenges and possibilities for future improvement of these models.

The first set of parameters needed for the floating oscillator models are the amide I site frequencies. Several maps correlating these frequencies with the electrostatic potential, as well as the electric field and electric field gradients found at or between the atoms in the amide bond have been presented in literature. The maps were constructed from electronic structure calculations on N-methyl acetamide (NMA), a model system for the peptide bond monomer. The chemical structure of NMA is shown in Fig. 1. All maps, except the Mukamel map, use the molecule deuterated on the nitrogen. With one exception, these maps assume a linear correlation between the frequency and the electrostatic parameters. The Mukamel map includes quadratic terms. We have selected the following set of maps for the present study.

Gradient. The Gradient map was obtained from DFT calculations on NMA embedded in electric charge environments. The electric field and electric field gradients at the C, O, N, and D atoms were used in this map. The map is transferable and reproduces the absorption and two-dimensional infrared spectra in several polar solvents (heavy water, acetonitrile, methanol-d, and DMSO-d$_6$).

Skinner. This map was constructed from DFT calculations on NMA surrounded by water molecules. The frequencies were correlated to the electric field at the C, O, N, and D atoms as generated by the TIP3P (Ref. 28) force field. The shape of the absorption spectrum of NMA in heavy water was well reproduced with this map, but the solvent shift was underestimated by about 25 cm$^{-1}$.

ChoK. This map was constructed using restricted Hartree-Fock (RHF) calculations on NMA surrounded by water molecules. The electrostatic potential at the C, O, N, and D atoms, as generated by the CHELPG force field water charges, was used in the parametrization. The line shapes for NMA in heavy water and DMSO-d$_6$ fit well with experiment.

The remaining maps were not included either because they did not consider the deuterated species, because they depend on the potential located on one or more C$_a$ atoms, or because they did not perform significantly different than those presented. The C$_a$ atoms are charged in one of the force fields that will be treated and this would result in diverging frequencies for the maps dependent on the potential at C$_a$. We will compare the site frequencies that are obtained using the partial charges from the GROMOS96 (Ref. 30) and OPLS (Ref. 31) force fields combined with the three maps described above.

FIG. 1. The chemical structures of N-methyl acetamide (NMA), N-acetyl-glycine-N'-methylamide (GLDP), and [Leu]-enkephalin. The amide bonds of GLDP and [Leu]-enkephalin are labeled with the site labels used in the text.
neighbor coupling and frequency shift model. Comparisons between different models predicting the GLDP and [Leu]-enkephalin spectra are presented in Sec. III. In Sec. IV some of the results will be discussed in more detail. Finally, conclusions are drawn in Sec. V. The Appendix gives an overview of the electrostatic coupling models.

II. CONSTRUCTION OF THE MODELS

DFT calculations were performed with the Amsterdam density functional (ADF) program. The ADF TZ2P basis was employed with the revised Perdew-Burke-Ernzerhof exchange correlation functional (RPBE). The NMA geometry was optimized and the normal modes and frequencies were obtained from the Hessian calculated numerically by distorting the structure. The gas phase frequency for the amide I mode of NMA is 1717 cm$^{-1}$. The DFT calculation underestimates it by a factor of 1.0255.

A transition charge coupling model, as described in the Appendix, was constructed by freezing the terminal methyl groups of NMA and calculating the normal modes with this restriction. The partial charges were obtained using the multipole derived charges. The charges on the terminal methyl groups were summed and placed on the methyl carbon. The charge flows were calculated using numerical displacements of the normal mode coordinate of 0.06 Å. The parameters are given in Table I and the expression for the TCC model is given in Eq. (A6). The new TCC model only differs from the original model by Hamm and Woutersen in the parameters. They used the Mulliken charges and normal modes from DFT calculations with the B3LYP exchange correlation potential.

The procedure used to calculate the normal modes of NMA was applied to GLDP. The Ramachandran angles between the two glycine units were fixed and all remaining coordinates were optimized. The Hessian was obtained for the optimized structure and the normal modes and their frequencies were found. The site frequencies and the coupling between them were generated using Hessian reconstruction. The magnitude of the CO stretch vibration was used to obtain the eigenvector matrices needed in the reconstruction. The frequencies and couplings were corrected by the factor of 1.0255 noted above. The frequency shifts of the two oscillators with respect to the NMA frequency were calculated by subtracting the NMA gas phase frequency. The calculations were performed for all configurations with the Ramachandran angles $0^\circ \leq \phi \leq 180^\circ$ and $-150^\circ \leq \psi \leq 180^\circ$ at 30$^\circ$ intervals for both angles. The configurations with negative $\phi$ are related by symmetry to those with positive values. The frequency shifts for the two oscillators and the coupling are shown in Fig. 2 as a function of the Ramachandran angles. The structure of these shift and coupling maps is identical to those of earlier studies.

The calculated couplings were used as a coupling map, denoted as the nearest-neighbor coupling (NNC) map.

We construct a nearest-neighbor frequency shift (NNFS) model for the site frequency based on combining the electrostatic maps with the frequency shifts in GLDP. In this model the site frequency of a given unit $i$ (numbering the units from the N-terminus) is given by

$$\omega_i = \omega_{\text{gas}} + \Delta\omega_N(\phi_{i-1, i}, \psi_{i-1, i}) + \Delta\omega_C(\phi_{i,i+1}, \psi_{i,i+1}) + \Delta\omega_{\text{map}}(\phi'(r), E'(r), \nabla E'(r)),$$

where $\omega_{\text{gas}}$ is the gas phase frequency (1717 cm$^{-1}$) and $\Delta\omega_N(\phi_{i-1, i}, \psi_{i-1, i})$ is the frequency shift calculated for the C site of GLDP under the influence of the N site. This shift depends on the Ramachandran angles between the $i$th unit and the $(i-1)$th unit. Similarly, $\Delta\omega_C(\phi_{i,i+1}, \psi_{i,i+1})$ is the frequency shift calculated for the N site of GLDP under the influence of the C site. This shift depends on the Ramachandran angles between the $i$th unit and the $(i+1)$th unit. $\Delta\omega_{\text{map}}(\phi'(r), E'(r), \nabla E'(r))$ is the shift caused by everything but the nearest neighbors. This depends on the electrostatic potential $\phi'(r)$, or field $E'(r)$ and gradient $\nabla E'(r)$, generated by all charges in the oscillator’s environment, excluding the charges in the peptide bond of unit $i$, $i-1$, and $i+1$. For the terminal groups, the electric field generated by the neighbor-

**TABLE I. Parameters for the transition charge model.** C(C) is the carbon atom of the methyl group bound to the amide carbon and C(N) is the carbon atom of the methyl group bound to the amide nitrogen. The normal mode coordinates $v$ are given in units of the amplitude 0.028 074 Å. (The x direction is along the CO bond, whereas the y axis is in the molecular plane pointing in the direction from C to N.)

<table>
<thead>
<tr>
<th>Atom</th>
<th>$q/e$</th>
<th>$dq/e$</th>
<th>$v_x$</th>
<th>$v_y$</th>
<th>$v_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(C)</td>
<td>0.11072</td>
<td>$-0.01668$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0.37173</td>
<td>0.02845</td>
<td>$-0.831$</td>
<td>0.105</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>$-0.48418$</td>
<td>$-0.01736$</td>
<td>0.074</td>
<td>$-0.036$</td>
<td>0</td>
</tr>
<tr>
<td>O</td>
<td>$-0.53632$</td>
<td>0.01530</td>
<td>0.517</td>
<td>$-0.047$</td>
<td>0</td>
</tr>
<tr>
<td>C(N)</td>
<td>0.29527</td>
<td>$-0.00963$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0.24278</td>
<td>$-0.00008$</td>
<td>0.073</td>
<td>$-0.133$</td>
<td>0</td>
</tr>
</tbody>
</table>
III. RESULTS

A. GLDP

The site frequencies were calculated using the different models for GLDP for all 24 combinations of the Ramachandran angles $0^\circ \leq \phi \leq 180^\circ$ and $-120^\circ \leq \psi \leq 120^\circ$ with $60^\circ$ intervals for both angles. The standard deviations between the model and DFT frequencies are given in Table II, along with the correlation coefficients. The best performance is seen for Cho4/OPLS. For all other maps the GROMOS force field gives smaller standard deviations than the OPLS force field. For the Gradient and Skinner maps with the OPLS force field, two groups of frequencies turn out to exist. The first group belongs to the N site and the frequencies in this group are on average about $8 \text{ cm}^{-1}$ above the DFT values. The second group belongs to the C site and the frequencies are about $40 \text{ cm}^{-1}$ too low. This is most likely because for the calculation of the frequencies for the C site a charge is located on the $C_\alpha$ between the two units. The distance between this charge and the carbon atom at the C site is only one bond length. The short distance leads to a relatively large error if the charge distribution deviates even slightly from the real charge distribution. Since the position of the $C_\alpha$ relative to the peptide bond is largely fixed, the error introduced in the C site frequency will be a systematic shift independent of the Ramachandran angles. The error is larger in the Gradient and Skinner maps than in the Cho4 map, since the latter depends on the electrostatic potential, which scales as the inverse distance, whereas the Gradient and Skinner maps depend on the square and cube of the inverse distance. If corrected by shifting the C site by the average error, the Gradient and Skinner maps perform as well as the Cho4 map for OPLS.

The coupling obtained from the DFT calculations was compared to that from the TCC model, the TDC model, and the Tasumi coupling map. The standard deviations, between these for the 24 GLDP configurations are 5.02, 7.63, and $6.55 \text{ cm}^{-1}$ for the TDC, TCC, and Tasumi maps, respectively. The TCC model in this case is in closest agreement with the DFT calculations. One would maybe expect to find a smaller deviation between the present DFT calculations and the Tasumi coupling map, which was obtained at the HF/6-31+G** level. In that study, however, only the structural parameters for the methylene group ($\text{CH}_2$) were optimized, while the rest of the structure was fixed to be identical to that of NMA. Furthermore, the normal mode coordinate of NMA was used, whereas we use the full normal modes of GLDP. This might lead to the larger deviations between our DFT calculations and the Tasumi coupling map. While the deviations for the TCC and TDC models are spread over all frequencies, the largest deviations for the Tasumi map occur for the configurations which are sterically hindered.

B. [Leu]-enkephalin

The site frequency models were tested on the small peptide [Leu]-enkephalin in the gas phase. The ten most stable configurations were used. Figure 4 shows the distribution of Ramachandran angles in the ten structures. The $\beta$-sheet-like Ramachandran angles ($\phi = -120^\circ$, $\psi = 120^\circ$)
dominate, but there are a few configurations in the opposite quadrant ($\phi \approx 120^\circ$, $\psi \approx -120^\circ$). In proteins, the latter are usually only found in connection with glycine. Glycine is more flexible than all the other amino acids because of the lack of substitution on the $C_n$. $\text{Leu}$-enkephalin contains two glycine units.

The frequencies obtained with the different frequency maps and force fields are shown against those from the DFT calculations as reported in Ref. 25 in the scatter plot in Fig. 5. The standard deviation and correlation coefficients between the models and the DFT frequencies are given in Table III. In Ref. 25 the frequencies were obtained using a model combining an electrostatic map with nearest-neighbor coupling and frequency shift maps which we refer to as the Hirst model. This model gives an overall standard deviation of 13.6 cm$^{-1}$ and a correlation of 0.51, which is comparable to the best of the models only based on electrostatic maps.

In Table IV, we compare the results obtained by combining the NNFS model introduced in Sec. II and Gradient/GROMOS for the long-range interactions (see Sec. II) with the Gradient/GROMOS, Gradient/OPLS, and Hirst models. Here the correlation between the DFT site frequency and the model frequency is given for each of the four sites separately (see Fig. 1 for labeling), along with the overall correlation.

![FIG. 5. Scatterplot of the model frequencies of $\text{Leu}$-enkephalin plotted against the frequency obtained from DFT calculations. The straight line illustrates where the points would be in case of perfect agreement with DFT.](image)

**TABLE III.** The standard deviation and correlation coefficient between the model site frequencies and the DFT frequencies for the set of ten $\text{Leu}$-enkephalin configurations described in the text.

<table>
<thead>
<tr>
<th>Force field</th>
<th>GROMOS</th>
<th>OPLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Map</td>
<td>Std. dev. (cm$^{-1}$)</td>
<td>Corr.</td>
</tr>
<tr>
<td>Gradient</td>
<td>12.0</td>
<td>0.55</td>
</tr>
<tr>
<td>Skinner</td>
<td>17.8</td>
<td>0.49</td>
</tr>
<tr>
<td>Cho4</td>
<td>11.8</td>
<td>0.64</td>
</tr>
</tbody>
</table>

![FIG. 6. Scatterplot of the model frequencies of $\text{Leu}$-enkephalin against the frequency from DFT calculations separated for the four sites. The straight lines illustrate where the points should ideally be. The arrows show the location of configuration 4. The circles are the Hirst model, the squares are the Gradient/GROMOS frequencies, the diamonds are Gradient/OPLS frequencies, and the crosses are the frequencies from the NNFS/Gradient/GROMOS method.](image)
models only consider the amide I vibrations and are therefore unlikely to give an accurate description of this configuration. Moreover, the accuracy of the DFT site frequencies obtained from the matrix reconstruction, neglecting the mixing of the amide I with the terminal carboxyl, should be used cautiously for this configuration. It is therefore reasonable to neglect configuration 4 when comparing the different models. The Gradient/OPLS and Hirst models both predict much too high frequencies on site 4. This is likely to be due to the charge parametrization on the terminal carbonyl group.

In the remainder of this section, we present two sets of results for the IR line spectra. In the first we use one fixed site frequency model and vary the coupling models. In the second we fix the coupling model and vary the site frequency model. The IR line spectra were calculated for all configurations using NNFS/Gradient/GROMOS for the site frequencies, the TCC, Tasumi, or NNC maps for the nearest-neighbor coupling, and the TDC or TCC models for the remaining couplings. The standard deviation \( \sigma = \sqrt{\langle (\omega_{\text{model}} - \omega_{\text{DFT}})^2 \rangle} \) between the obtained eigenfrequencies and the DFT results in Ref. 25, excluding configuration 4, is reported in Table V, along with the correlation of the intensities and the standard deviation from the average error \( \sigma' = \sqrt{\langle (\omega_{\text{model}} - \omega_{\text{DFT}} - \langle \Delta \omega_{\text{err}} \rangle)^2 \rangle} \), where the average error is \( \langle \Delta \omega_{\text{err}} \rangle = \langle \omega_{\text{model}} - \omega_{\text{DFT}} \rangle \). Here \( \langle \ldots \rangle \) denotes the average over all site frequencies in the nine configurations. When the Tasumi or NNC map is used for the nearest-neighbor coupling, a clear improvement of the intensity correlation is seen. For the long-range interactions the TCC model is slightly better than the TDC model, while for the nearest-neighbor coupling the NNC map is slightly better than the Tasumi map. The standard deviation is rather insensitive to the change of coupling model and in all cases is roughly 7 cm\(^{-1}\). Also the standard deviation from the average error is rather insensitive to the coupling model, but much smaller. This indicates that the errors in the eigenfrequencies are rather systematic. After shifting the spectrum by the average error \( \langle \Delta \omega_{\text{err}} \rangle \), the standard deviations are just 3.5 cm\(^{-1}\).

The IR line spectra were calculated using the coupling model performing best in the previous paragraph (the NNC map and the TCC model for long-range couplings) and varying the site frequency models. For the latter, five different choices were considered, listed at the bottom of Table V, which for each model gives the standard deviation, intensity correlation, and standard deviation from the average error. The standard deviations for the electrostatic map models are \( \sim 1 \) cm\(^{-1}\) smaller than for the NNFS based models. The deviation is, however, less systematic and the standard deviation from the average error is \( \sim 1-2 \) cm\(^{-1}\) larger than for the NNFS based models. The intensity correlations for the electrostatic map models are lower than for the NNFS based models as well. For the NNFS based models the spectral features are well reproduced, but the model spectrum is blue-shifted by about 6 cm\(^{-1}\). The difference in accuracy between the NNFS/Cho4/GROMOS and NNFS/Gradient/GROMOS is very small, with the latter model agreeing slightly better with the full DFT calculation. Spectra for the best model are plotted for all ten configurations in Fig. 7. In these spectra the features are well reproduced with the exception of configuration 4 that has already been discussed.

![Normalized IR line spectra for the ten configurations of [Leu]-enkephalin. The full lines are DFT absorption lines taken from Ref. 25. The striped lines are the absorption lines obtained with the NNFS/Gradient/GROMOS/TCC/NNC model explained in the text.](image)

TABLE V. Eigenfrequency deviations and intensity correlations for nine enkephalin configurations with different models compared to DFT calculations. The standard deviations (\( \sigma \)) and standard deviations from the average error (\( \sigma' \)) are given in cm\(^{-1}\).

<table>
<thead>
<tr>
<th>Model</th>
<th>Charges</th>
<th>Coupling</th>
<th>( \sigma )</th>
<th>( \sigma' )</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TDC/TCC</td>
<td>7.0</td>
<td>3.1</td>
<td>0.68</td>
</tr>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TCC/TCC</td>
<td>7.0</td>
<td>3.0</td>
<td>0.70</td>
</tr>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TDC/Tasumi</td>
<td>7.3</td>
<td>3.8</td>
<td>0.93</td>
</tr>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TCC/Tasumi</td>
<td>7.3</td>
<td>3.8</td>
<td>0.94</td>
</tr>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TDC/NNC</td>
<td>7.0</td>
<td>3.3</td>
<td>0.93</td>
</tr>
<tr>
<td>NNFS/Gradient</td>
<td>GROMOS</td>
<td>TCC/NNC</td>
<td>7.1</td>
<td>3.4</td>
<td>0.94</td>
</tr>
<tr>
<td>NNFS/Cho4</td>
<td>GROMOS</td>
<td>TCC/NNC</td>
<td>7.5</td>
<td>3.6</td>
<td>0.94</td>
</tr>
<tr>
<td>Cho4</td>
<td>GROMOS</td>
<td>TCC/NNC</td>
<td>6.0</td>
<td>5.3</td>
<td>0.90</td>
</tr>
<tr>
<td>Cho4</td>
<td>OPLS</td>
<td>TCC/NNC</td>
<td>6.1</td>
<td>4.6</td>
<td>0.89</td>
</tr>
<tr>
<td>Gradient</td>
<td>GROMOS</td>
<td>TCC/NNC</td>
<td>5.8</td>
<td>5.6</td>
<td>0.83</td>
</tr>
</tbody>
</table>
IV. DISCUSSION

In the previous section a few shortcomings of the coupling and frequency models and differences between them were found. In the following these issues are discussed.

We have seen that for GLDP the TCC model agrees better with our DFT calculations than the TDC and Tasumi models. In [Leu]-enkephalin both the Tasumi map and the new NNC map gave better results than the TCC model. One possible reason for this discrepancy is that in [Leu]-enkephalin the Ramachandran angles are predominantly β-sheet-like, while in the GLDP calculations all Ramachandran angles were equally sampled. It turns out that for β-sheet-like configurations the Tasumi map agrees well with our DFT calculations, while large discrepancies exist for Ramachandran angles which result in sterical hindrance. The difference between the NNC and Tasumi maps shows the importance of structural relaxation of the sterically hindered configurations, which is neglected in the latter.

The Hirst and Gradient/OPLS site frequency models both overestimate the frequency of site 4 significantly. Most likely this is due to the charges on the terminal carboxyl group directly attached to site 4. The small spatial separation makes the terminal sites extra sensitive to the charge parametrization on the terminus. This would only pose a minor problem in large proteins, where this affects a small fraction of the sites. Moreover, maps accounting for the frequency shifts generated by the most common termini would solve this problem entirely.

The predicted eigenfrequencies of [Leu]-enkephalin were about 6 cm$^{-1}$ too high. This systematic peak shift could be due to the use of different exchange correlation functionals in the DFT calculations used to construct the NNFS and Gradient map and the one used to calculate the [Leu]-enkephalin frequencies. Coupling to backbone and side chain modes can give rise to differences between the floating oscillator models and the full DFT calculation as well, but these are likely to depend on the geometry and thus to be less systematic. Shortcomings of the electrostatic map, such as the use of atomic point charges, could contribute to the systematic error as well. However, again such errors are likely to be geometry dependent and thus less systematic.

For one of the [Leu]-enkephalin configurations none of the models worked. This particular configuration contains a hydrogen bond between the C-terminus and the N-terminus causing the amide I modes to mix with a mode on the C-terminus. In order to treat this configuration one would need to include the mode on the C-terminus in the floating oscillator model. The probability of encountering such configurations in other peptides or proteins is small, but not negligible. It is, in most cases, acceptable to neglect the vibration of the terminal acid groups and side chain acid groups. However, one can imagine that in specific proteins the terminal or side chain acid groups form hydrogen bonds to the backbone and that this gives rise to small frequency differences and strong couplings. In order to treat such systems it is necessary to develop similar maps for the acid groups as those developed for the amide groups. Certain amino acid residues contain vibrations close to the amide I region. One should be cautious when treating these systems, since coupling with such side chain vibrations can affect the amide I spectrum.

V. CONCLUSION

In this paper, we studied the predictive power of floating oscillator models for the amide I band of polypeptides. These models are characterized by the eigenfrequencies of the oscillators and the interactions between them. Both sets of parameters are affected by the conformation of the polypeptide as well as its environment (the solvent); here we focused on the former effect, by restricting ourselves to polypeptides in the gas phase. We introduced new models that relate the site frequencies and the interactions to the conformation and assessed these and other, existing models, by comparison to DFT calculations for [Leu]-enkephalin, which contains four amide groups.

We introduced the following new elements in the parametrization of the oscillator model: (i) A new TCC model, which introduces multipole effects through atomic point charges and transition charges, (ii) a new map which relates the nearest-neighbor interactions to the Ramachandran angles (this incorporates through-bond coupling as well as polarization effects), and (iii) a new model for the shift in the site frequencies generated by the covalently bound neighboring units (this accounts more accurately for the electrostatic shift due to the use of the DFT charge density instead of fixed atomic point charges and the model includes polarization effects as well). The models for (ii) and (iii) were constructed by using DFT calculations on the dipeptide GLDP. Previous model parametrizations that we considered for comparison used, amongst others, the Tasumi map or the TDC model for the interactions and the purely electrostatic maps for the shift in the site frequencies.

The new TCC model improves the description of the long-range interactions. The effect on the [Leu]-enkephalin spectra was, however, not as pronounced as the improvement gained with the new nearest-neighbor map for the short-range interactions. This shows that the multipole contribution is less important for the spectra than the nearest-neighbor effects and that through-bond coupling is crucial. As a result, for the nearest-neighbor coupling it is necessary to use a map from electronic structure calculations. The new Ramachandran angle dependent DFT map for the nearest-neighbor interactions is significantly better than the purely electrostatic maps that at best have an accuracy of ±10 cm$^{-1}$ in both GLDP and [Leu]-enkephalin. This shows the importance of an accurate description of the electron density for the short-range electronic interactions and of possible contributions from polarization effects.

The present study shows that the use of nearest-neighbor maps for both the coupling and the frequency shift improves the predicted spectra significantly. From our calculations for [Leu]-enkephalin, we concluded that the new model including long-range transition charge coupling and a DFT map for nearest-neighbor interactions performs best. Corrected for a systematic blueshift an error of only 3.5 cm$^{-1}$ is found for the frequencies. The remaining deviations may be due to
many factors. Dispersion forces are neglected, the charges in the force fields (especially for the side chains) do not give an exact description of the real charge distribution, and the effects of terminal amine and acid groups were neglected for the terminal amide vibrations. The accuracy found here, however, will be adequate for most purposes. The predicted spectral structure for the different configurations is recognizable and the proposed models should be good enough to predict and interpret peptide spectra. In future work we plan to investigate how the models presented perform for peptides in solution.

ACKNOWLEDGMENTS

One of the authors (T.I.C.J.) acknowledges the Netherlands Organization for Scientific Research (NWO) for support through a VENI grant. The authors thank Andrei Tokmakoff and Ziad Ganim for helpful discussions.

APPENDIX: COUPLING MODELS

Assuming that the coupling is due to electrostatic interactions between the oscillators, it is determined by the derivative of the electrostatic interaction energy between the two oscillators with respect to their coordinates \(x_i\) and \(x_j\),

\[
J_{ij} = \frac{1}{4\pi\varepsilon_0} \left[ \frac{\partial^2}{\partial x_k \partial x_l} \int \frac{\rho_k(x_i, r_k) \rho_l(x_j, r_l)}{|r_i - r_j|} dr_k dr_l \right]_{x_i = 0, x_j = 0}.
\]  

(A1)

Here \(\rho_k(x_i, r_i)\) is the charge density in the vicinity of oscillator \(i\) at the point \(r_i\) when the oscillator is displaced \(x_i\) along the normal mode coordinate with respect to its equilibrium position. Displacing the coordinate by \(x_i\) corresponds to displacing the involved nuclei by \(x_i v_{n,i}\), where \(v_{n,i}\) is the normal mode coordinate for nucleus \(n\) of oscillator \(i\) and \(x_i\) is a dimensionless quantity. The charge densities at the two oscillators, \(\rho_i\) and \(\rho_j\), are taken to be nonoverlapping and the charge density in one unit is taken to be independent of nuclear displacements in the other one.

If the oscillators are separated by distances much larger than their size, the electrostatic interaction can safely be approximated by its dipole-dipole part. This gives rise to the transition dipole coupling model (TDC),

\[
J_{ij} = \frac{1}{4\pi\varepsilon_0} \left[ \frac{\mu_i \cdot \mu_j}{r_{ij}^3} - 3 \left( \frac{\mu_i \cdot r_{ij}}{r_{ij}^5} \right) \left( \frac{\mu_j \cdot r_{ij}}{r_{ij}^5} \right) \right].
\]  

(A2)

The transition dipole is given by the derivative of the dipole with respect to the displacement \(x_i\) at the equilibrium position,

\[
\mu_i = \left[ \frac{\partial}{\partial x_i} \left( \int \rho_i(x, r) r dr \right) \right]_{x_i = 0}.
\]  

(A3)

If the distance between the oscillators is of the same order of magnitude as their size, one needs to account for the multipole interactions as well. In order to simplify the calculation of the integral over the charge densities, the latter can be approximated by point charges. Although, in principle, these charges can be located anywhere suitable, we position them at the atomic sites. The charge density then becomes

\[
\rho_i(x, r) = \sum_n \delta(r - r_n(x)) (q_n + dq_n x),
\]  

(A4)

where \(r_n(x)\) is the position of atom \(n\) given the oscillator displacement \(x_n, q_n\) is the partial charge at atom \(n\), and \(dq_n\) is the derivative of the partial charge with respect to the oscillator coordinate. In principle, the partial charge does not need to be linear in this coordinate, but the higher-order derivatives will not contribute to the coupling in Eq. (A1) and are therefore neglected. Inserting the point charge expression in Eq. (A1) leads to

\[
J_{ij} = \frac{1}{4\pi\varepsilon_0} \left[ \frac{\partial}{\partial x_k} \sum_{n,m} \left( q_n + dq_n x_n \right) \left( q_m + dq_m x_m \right) \right]_{x_i = 0, x_j = 0}.
\]  

(A5)

Using the fact that the coordinate of point charge \(n\) is given by \(r_n(x) = r_n(0) + x_n v_{n,i}\), the coupling can be calculated analytically from the above equation. The full expression for the transition charge coupling is

\[
J_{ij} = \frac{1}{4\pi\varepsilon_0} \sum_{n,m} \left( \frac{dq_n dq_m}{|r_{n,m}|^3} \left( \frac{v_{n,i} \cdot r_{n,m}}{|r_{n,m}|^5} \right) - \frac{dq_n dq_m v_{n,i} \cdot r_{n,m}}{|r_{n,m}|^5} - \frac{dq_n dq_m v_{n,i} \cdot r_{n,m}}{|r_{n,m}|^5} \right).
\]  

(A6)

The vector \(r_{n,m}\) is a short hand notation for the relative position vector \(r_n(x) - r_m(x)\).

The partial charges can be found using Mulliken population analysis45 and the charge flows \(dq_n\) are obtained by numerical differentiation of the partial charges with respect to the vibrational coordinate.35 Alternatively, these parameters can be obtained using multipole derived charges (MDC),36 which is a method that provides partial charges reproducing the molecular multipoles as obtained in the electronic structure calculation up to a desired level. The latter method is used in the current paper.

When two oscillators are so close that their charge densities overlap, these coupling models all break down. In this case through-bond effects can be expected to affect the coupling. For nearest neighbors in a peptide chain the charge densities can be expected to overlap and the through-bond
effects need to be addressed. This is best done by performing electronic structure calculations on dimers as we do in Sec. II. The two Ramachandran angles $\phi$ and $\psi$ determine the structure of the peptide backbone and can be used to create coupling maps from dimer configurations.

46 See EPAPS Document No. E-JCPSA6-125-016627 for the nearest-neighbor frequency shift and nearest-neighbor coupling data that are shown in Fig. 2 and used in the map in Eq. (2). This document can be reached via a direct link in the online article’s HTML reference section or via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html).