

## Supplementary Information

# Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field

Florian Thauhay,<sup>a</sup> Jean-Pierre Dognon,<sup>b</sup> Gilles Ohanessian,<sup>a</sup> and Carine Clavaguéra<sup>\*a</sup>

<sup>a</sup> *Laboratoire de Chimie Moléculaire, Ecole polytechnique, CNRS, 91128 Palaiseau Cedex, France; E-mail: carine.clavaguera@polytechnique.edu*

<sup>b</sup> *CEA/Saclay, DSM/IRAMIS/NIMBE, CNRS, Laboratoire de Chimie Moléculaire et de Catalyse pour l'Energie, 91191 Gif-sur-Yvette, France.*

### 1 Low-energy structure determination

The Replica Exchange Molecular Dynamics method (REMD) was used to explore the complex conformational landscape of the Ac-Phe-Ala-NH<sub>2</sub> dipeptide. The coupling of REMD with a polarizable force field such AMOEBA is a powerful tool for an efficient sampling of the potential energy surface of flexible systems in order to locate low-energy conformations as discussed in previous studies<sup>1,2</sup>. Briefly, it consists in running several MD trajectories at various temperatures, and punctually attempting an exchange between two configurations of neighboring trajectories. This exchange is accepted with a Metropolis probability. In the present REMD simulations, 12 replica were performed using the AMOEBA force field. The temperature range was chosen with a geometric progression between 200 and 600 K, that allows both an effective exploration of minima and barrier crossing between potential energy basins. The REMD is coupled with TINKER via our own adaptation of TiReX software<sup>1,3</sup>. Each trajectory was propagated with a time step of 1 fs, and the temperature was controlled with a Nosé-Hoover thermostat. Exchanges of a random pair of adjacent configurations were attempted every 10 ps and configurations were saved every 2 ps. The total simulation time for each replica was 2 ns. At the end, all resulting configurations were optimized with AMOEBA and were sorted to eliminate duplicates. Finally, 10 low-energy structures have been selected for quantum chemistry calculations. Three functionals were tested for geometry optimization, i.e. M06, B97-D and B3LYP, associated to the 6-311+G(d,p) basis set. At the DFT/M06 optimized structures, single point energy calculations were performed with the Dunning's correlation consistent cc-pVTZ basis set and the same functional. The geometries obtained with the three functionals are very similar and the B3LYP/6-311+G(d,p) energetic order was chosen for the nomenclature. The high flexibility of the peptide allows various intramolecular non-covalent interactions. These interactions often correspond to hydrogen bonds between an amine or amide groupe and a neighboring oxygen in a C<sub>5</sub> or C<sub>7</sub> motif. Furthermore, C<sub>8</sub> and C<sub>10</sub> interactions are possible between more distant groups. Two kinds of C<sub>7</sub>B interactions exist: in FAd, the interaction occurs in a regular C<sub>7</sub> cycle, whereas the cycle is distorted in FAF. It was shown previously on several structures with different C<sub>7</sub> distances that the hydrogen bond strength is not the only factor to determine the stability of a structure<sup>4</sup>. The following section will describe the change in conformation during MD simulations and one can see that even if this C<sub>7</sub> cycle is distorted, it is still relevant to stabilize the structure. Another stabilizing effect is the  $\pi$  interaction that can exist between an amide hydrogen and the aromatic ring.

The relative energies of the 10 confirmations were computed at four levels of calculation (Table S2).

The energetic order can be impacted by the functional used, however the differences are rather small and all methods yield FAa as the most stable structure, similarly than in previous studies<sup>5,6</sup>. This conformation is stabilized by a C<sub>5</sub>, a C<sub>7</sub>B and a H<sub>ala</sub>- $\pi$  interaction.

The main objective of this paper is the normal mode analysis of finite temperature spectra. This implies that the followed structures remain stable during the dynamics at a given temperature. At 50 K and even more at 200 K, barriers between conformations are easier to be crossed. Indeed, Table S1 illustrates the reorganization of some structures during the simulations used to record the IR spectra. From the ten starting structures, six do not keep their original conformation at 50 K, and seven at 200 K. Reorganizations during the dynamics lead to the same stable conformations at both temperatures. At 50 K, with two multipole sets, six structures change to another conformation from M06/6-311+G(d,p) multipoles, and five from MP2/cc-pVDZ, and the remaining stable conformations are the same at both levels. Finally, the choice of the multipole set is not negligible and it will be discussed in the following section. As the structures FAa, FAd, FAe, FAF and FAh do not reorganize with the multipoles extracted at MP2/cc-pVDZ level of theory, the following assignments of spectra will be discussed only for these structures.

**Table S1** Structure reorganizations during a 600 ps simulation using 2 multipole sets and 2 temperatures. Only the structures that reorganized are presented.

| M06 multipoles                                      |   | MP2 multipoles                                      |
|---|---|---|
| 50 K  | 200 K   | 50 K  |
| FAb $\rightarrow$ FAF                               | FAb $\rightarrow$ FAF                               | FAb $\rightarrow$ FAF                               |
| FAc : C <sub>8</sub> $\rightarrow$ C <sub>7</sub> B | FAc : C <sub>8</sub> $\rightarrow$ C <sub>7</sub> B | FAc : C <sub>8</sub> $\rightarrow$ C <sub>7</sub> B |
|   | FAe $\rightarrow$ FAF                               |   |
| FAG $\rightarrow$ FAa                               | FAG $\rightarrow$ FAa                               | FAG $\rightarrow$ FAa                               |
| FAh $\rightarrow$ FAF                               | FAh $\rightarrow$ FAF                               |   |
| FAi $\rightarrow$ (mirror) FAe                      | FAi $\rightarrow$ (mirror) FAe                      | FAi $\rightarrow$ (mirror) FAe                      |
| FAj $\rightarrow$ FAF                               | FAj $\rightarrow$ FAF                               | FAj $\rightarrow$ FAF                               |

**Table S2** Relative energies in kcal.mol<sup>-1</sup> for the 10 low-energy conformations of FA using various functionals and basis sets.

|     | <sup>a</sup> B3LYP | <sup>a</sup> B97-D | <sup>a</sup> M06 | <sup>b</sup> M06 |
|-----|--------------------|--------------------|------------------|------------------|
| FAa | 0.00               | 0.00               | 0.00             | 0.00             |
| FAb | 1.02               | 2.11               | 2.67             | 2.74             |
| FAc | 1.26               | 1.32               | 5.38             | 5.47             |
| FAd | 1.27               | 1.54               | 2.79             | 2.67             |
| FAe | 1.91               | 1.54               | 1.65             | 1.48             |
| FAf | 1.95               | 1.75               | 3.35             | 3.32             |
| FAG | 2.15               | 3.61               | 2.97             | 2.68             |
| FAh | 2.36               | 4.00               | 4.03             | 3.84             |
| FAi | 5.68               | 5.59               | 3.31             | 2.93             |
| FAj | 9.84               | 2.35               | 6.70             | 6.60             |

<sup>a</sup>: 6-311+G(d,p), <sup>b</sup>: cc-pVTZ single point at M06/6-311+G(d,p) geometry

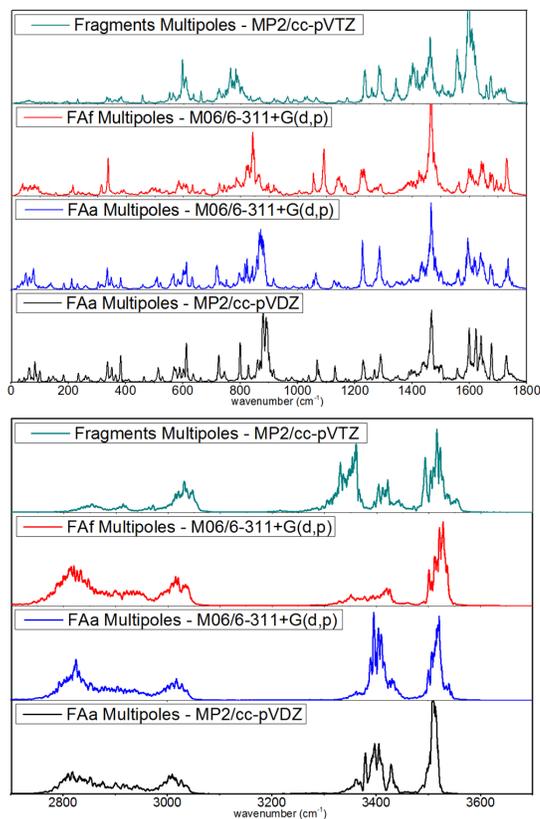
## 2 Choice of the Multipole set

The validation of a set of multipoles can be a long task depending on the target properties. Our purpose here is to generate low-energy structures and to reproduce IR spectra using both quantum chemistry and experimental data for the validation. For the dipeptide, the procedure was based on an iterative approach which consists in a generation of structures using REMD exploration, then DFT or MP2 geometry optimization in order to extract multipoles at the same level, and finally calculation of finite temperature MD spectra. Multipoles were derived from M06/6-311++G(d,p) and MP2/cc-pVDZ or cc-pVTZ calculations using the Gaussian<sup>7</sup> and GDMA<sup>8</sup> program packages.

The first set of multipoles was generated from two separate parts, i.e. the alanine (Ace-Ala-NH<sub>2</sub>) and phenylalanine (Ace-Phe-NH<sub>2</sub>) fragments, both protected, from the MP2/cc-pVTZ wave function. However, it was not possible to keep this level to extract multipoles on the whole dipeptide. Two levels of theory were thus chosen: MP2/cc-pVDZ and M06/6-311++G(d,p). The M06 functional was selected for its performance in reproducing non-covalent interactions<sup>9</sup>. At the M06 level, the iterative procedure led to FAa as the most stable structure. Furthermore, as this structure remained stable during MD, a multipole extraction was also performed using the MP2 wave function. The effects on the IR spectrum of FAa are presented in Fig. S1 by comparing diverse multipole sets, i.e. extracted from separate fragments, from the FAF or FAa structure at the M06 level, from FAa at the MP2 level. The comparison of the computed spectra shows their sensitivity to the quantum chemistry method used to obtain the multipole set. Some patterns of the spectrum are more impacted than others, such as the N-H stretch range (3300-3600 cm<sup>-1</sup>), the C=O stretch modes (1600-1800 cm<sup>-1</sup>), or the amide bands (1300-1500 cm<sup>-1</sup>). For these modes, frequency shifts occur due to their subtle response to intramolecular hydrogen bonds.

For example in the NH stretching range, three bands are present in the spectrum computed with multipoles obtained from fragments in comparison with only two bands for the other spectra computed with multipoles extracted on the whole peptide. Furthermore, the red-shift of this third band can be attributed to an extraction on separate fragments without taking into account hydrogen bonds between the two residues. Similar effects are observed on C=O stretch modes. The effect coming from the peptide conformation, FAa vs. FAF, used for the extraction can be compared on Fig. S1 at the same level of theory (red and blue curves). Again, the N-H stretch range is the most affected due to the memory of hydrogen bonds in the reference conformation. When going from DFT/M06 to MP2 levels, the differences are less significant. On N-H stretch frequencies, the average difference between experimental<sup>5</sup> and computed (T = 50 K) frequencies is less than 6 cm<sup>-1</sup> using both multipole sets. Regarding intensities, the C-H stretch pattern is affected when using multipole extracted from fragments because the lack of NH- $\pi$  interaction in the reference structure may induce a lower polarisation of the carbon and hydrogen atoms of the ring. Intensities in the N-H range are also impacted for similar reasons.

We conclude on the importance of including non-covalent interactions in multipole extraction which must be coupled to a high-level quantum chemistry method. In the case of FA, both M06 and MP2 multipoles can be used to perform DACF spectra. To be consistent with the general framework of AMOEBA, the spectra presented in our study are performed with the multipole set extracted from FAa structure at MP2/cc-pVDZ level. From the DMA procedure, a set of multipoles (charge, dipole and quadrupole) is assigned on each atomic site and multipoles are averaged over equivalent atoms, except for NH<sub>ala</sub> and NH<sub>phe</sub>, to be able to reproduce the difference of the environment between FAa and FAF structures.

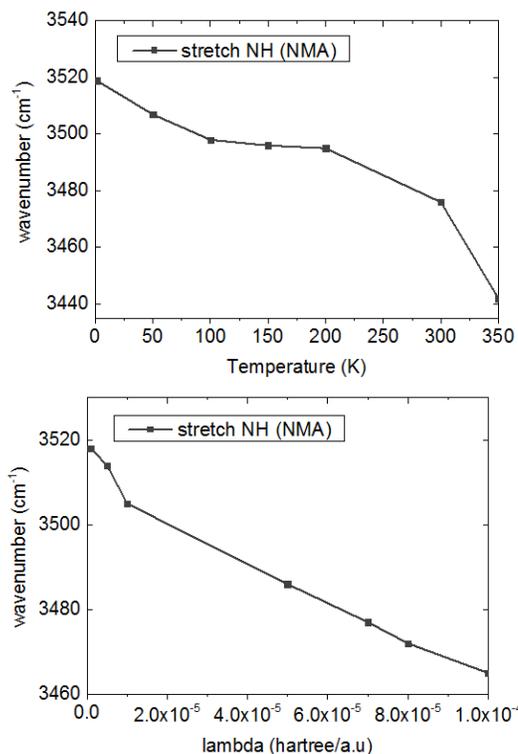


**Fig. S1** DACF-AMOEBA absorption spectra of FAa with 4 multipole sets in two spectral range. From top to bottom: multipoles extracted from protected alanine and phenylalanine separate fragments at the MP2/cc-pVTZ level (green), multipoles extracted from the FAf structure at the M06/6-311++G(d,p) level (red), multipoles extracted from the FAa structure at M06/6-311++G(d,p) level (blue), multipoles extracted from FAa at MP2/cc-pVDZ level (black). Spectra are recorded 300 ps at 200 K.

### 3 DMD based tools for vibrational mode analysis

The numerical accuracies on the internal coordinates given by TINKER software was increased from  $10^{-6}$  for cartesian coordinates,  $10^{-5}$  for bond length and  $10^{-4}$  for angles, to  $10^{-8}$  for cartesian coordinates,  $10^{-7}$  for bond length and  $10^{-6}$  for angles (DIGITS keyword).

In Fig. S2, the position of the maximum intensity of N-H stretch (NMA molecule) is collected from both simulations. The idea is to calibrate the amplitude of the driving force to reproduce the observed shifts with temperature. For example, at 200 K, the maximum amplitude of N-H stretch is observed close to  $3495 \text{ cm}^{-1}$ .



**Fig. S2** Shift of NH stretching mode in NMA as a function of the temperature or absorbed energy. Top: 100 ps of DACF simulation in the 0-350 K range. Bottom: each point corresponds to a 20 ps DMD simulation in the  $3450\text{-}3550 \text{ cm}^{-1}$  range for different  $\lambda$  values. At 200 K, the maximum amplitude of N-H stretch is observed close to  $3495 \text{ cm}^{-1}$ . This shift can be reproduced with a 20 ps DMD simulation with  $\lambda = 3.10^{-5}$  Hartree/Bohr.

#### 4 Technical details of DMD attributions for the Ac-Phe-Ala-NH<sub>2</sub> dipeptide

The structure was first energy minimized with TINKER software until the RMS gradient was less than  $10^{-5}$  kcal.mol<sup>-1</sup>, to avoid a more or less long period, in which the resonant coordinate hardly absorbs energy, which is distributed on other internal coordinates. After optimization, with only 10 000 steps, the emergence of a resonance was clearly observed. It is generally recommended to use a maximum integration step of 1/10 of the period of the signal to be observed (0.8 fs at 4000 cm<sup>-1</sup> for example). The mode analysis tools are based on a precise calculation of internal coordinates oscillation frequency, so we have to increase slightly the accuracies of this criterion : 0.1 fs in the 1000-4000 cm<sup>-1</sup> range and 0.2 fs below. The parameter  $\lambda$  manages the amplitude of the driving force and has to be chosen accurately. It should allow rapid excitation modes while guaranteeing an anharmonicity close to that given by the temperature of the DACF spectrum to be analyzed. For 1 K DACF spectrum assignments, harmonic oscillations are searched.  $\lambda$  was set to  $10^{-6}$  Hartree/Bohr in the 1000-4000 cm<sup>-1</sup> range and  $5.10^{-7}$  Hartree/Bohr below, because modes are softer on average. At 200 K, temperature allows numerous couplings between modes and more anharmonic behaviours. To aim these modes, we adjust lambda at  $2.10^{-5}$  Hartree/Bohr in the 1000-4000 cm<sup>-1</sup> range and  $1.10^{-5}$  Hartree/Bohr below.

## References

- 1 D. Semrouni, O. Balaj, F. Calvo, C. Correia, C. Clavaguéra and G. Ohanessian, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 728–738.
- 2 T. D. Rasmussen, P. Ren, J. W. Ponder and F. Jensen, *Int. J. Quantum Chem.*, 2007, **107**, 1390–1395.
- 3 E. S. Penev, S. Lampoudi and J.-E. Shea, *Comput. Phys. Commun.*, 2009, **180**, 2013–2019.
- 4 W. Chin, *PhD thesis*, Conformational landscape of small model peptides in the gas phase : UV/IR spectroscopy and theoretical approach, Paris Sud University, France, 2005.
- 5 W. Chin, F. Piuze, J.-P. Dognon, I. Dimicoli and M. Mons, *J. Chem. Phys.*, 2005, **123**, 084301.
- 6 S. Jaque, J. Oomens, A. Cimas, M.-P. Gaigeot and A. M. Rijs, *Angew. Chem. Int. Ed.*, 2014, **53**, 3663–3666.
- 7 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.
- 8 A. J. Stone, *Distributed Multipole Analysis for Gaussian Wavefunctions*, version 2.2.03, 2005-2007.
- 9 Y. Zhao and D. Truhlar, *Theo. Chem. Acc.*, 2008, **120**, 215–241.