Supplementary material for: The Vibrational Spectrum of the hydrated Alanine-Leucine Peptide in the Amide region from IR experiments and First Principles Calculation

Irtaza Hassan,[†] Luca Donati,[‡] Till Stensitzki,[†] Bettina G. Keller,[‡] Karsten Heyne,[†] and Petra Imhof[*,†]

†*Institute of Theoretical Physics, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany*

‡*Institute of Chemistry and Biochemistry, Freie Universität Berlin, Takustr. 3, 14195 Berlin, Germany*

E-mail: petra.imhof@fu-berlin.de

Supplementary Information

Molecular mechanics simulations

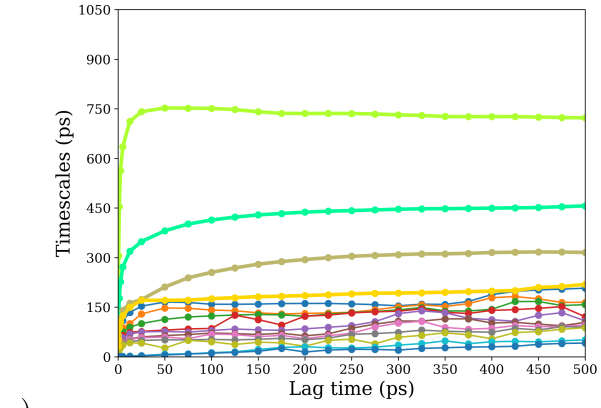
We performed classical molecular dynamics (MD) simulations of the Ala-Leu peptide in a cubic simulation box of explicit water (TIP3P water model[1]) employing the AMBER ff99SB-ILDN[2,3] force field. We used a minimum distance of 1 nm between the solute and the periodic boundaries of the box. Water hydrogen atoms and polar hydrogen atoms of the peptide (ND3 and ND) were modelled with the mass of deuterium, mimicking the experimentally measured system. For Lennard-Jones interactions and electrostatic interactions (Particle-Mesh Ewald[4,5] with grid spacing of 0.16 and an interpolation order of 4) we used cut-off value of 1 nm.

The system was minimised and equilibrated for 500 ps. Then three MD simulations of 400 ns each were launched which yields a total simulation time of 1.2 μs. A V-rescale thermostat[6] was applied to control the temperature at 300 K (NVT ensemble). The positions of the solute atoms were saved to file every 0.25 ps. No constraints were applied and the leap-frog integrator with the time step of 1 fs was employed using the GROMACS simulation package[7].

Markov state model

A transition matrix has been computed for transitions between the conformational clusters in the classical molecular dynamics trajectory with varying the lag time τ up to 500ps. The spectrum of the matrix, calculated for these lag times, indicates a convergence of the implied time scales (see equation 3) of the slowest process at about 50 ps. The implied time time-scales in Figure S1 suggest three slow processes, corresponding to transitions between four meta-stable sets. The spectral gap of the transition matrix is, however, after the third eigenvalue (see Table in Figure S1 b)). Therefore, we tried to perform PCCA+ to identify three and four meta-stable sets,

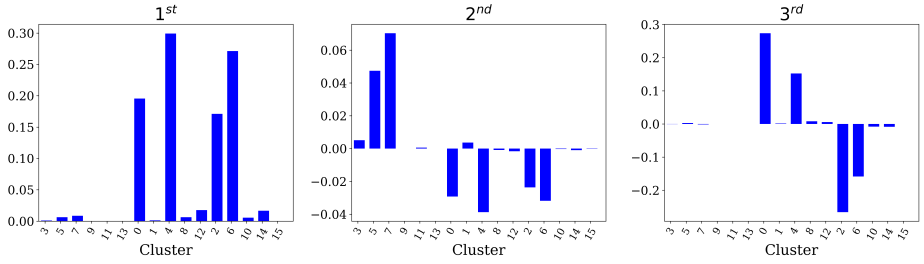
respectively. The grouping of micro-states into the meta-stable sets is listed in Table S1. The fourth partitioning results in a regrouping of formerly separated micro-states, denoting that there are several transitions on this time-scale, as already suggested by the eigenvalues. We therefore worked with a MSM of three meta-stable sets as presented in Figure 3. The first three eigenvectors are given in (see Figure S1 b)). The first eigenvector corresponds to the stationary distribution and the other two correspond to the slowest processes which can be understood as transitions between meta-stable sets of clusters.



a)

b)

Eigenvalues
1.000
0.936
0.877
0.789
0.746
0.739



c)

Figure S1: a) Implied time scales, b) First eigenvalues of the transitions matrix sampled with lag time τ=50ps. c) Corresponding first three eigenvectors expressed as contributions of the 16 micro-states.

Table S1: Meta-stable sets of micro-states according to PCCA+

I	II		
[3, 5, 7, 9, 11, 13]	[0, 1, 2, 4, 6, 8, 10, 12, 14, 15]		
I	II	III	
[3, 5, 7, 9, 11, 13]	[0, 1, 4, 8, 12]	[2, 6, 10, 14, 15]	
I	II	III	IV
[5, 7, 13]	[0, 1, 8]	[4, 6, 12, 14]	[2, 3, 9, 10, 11, 15]

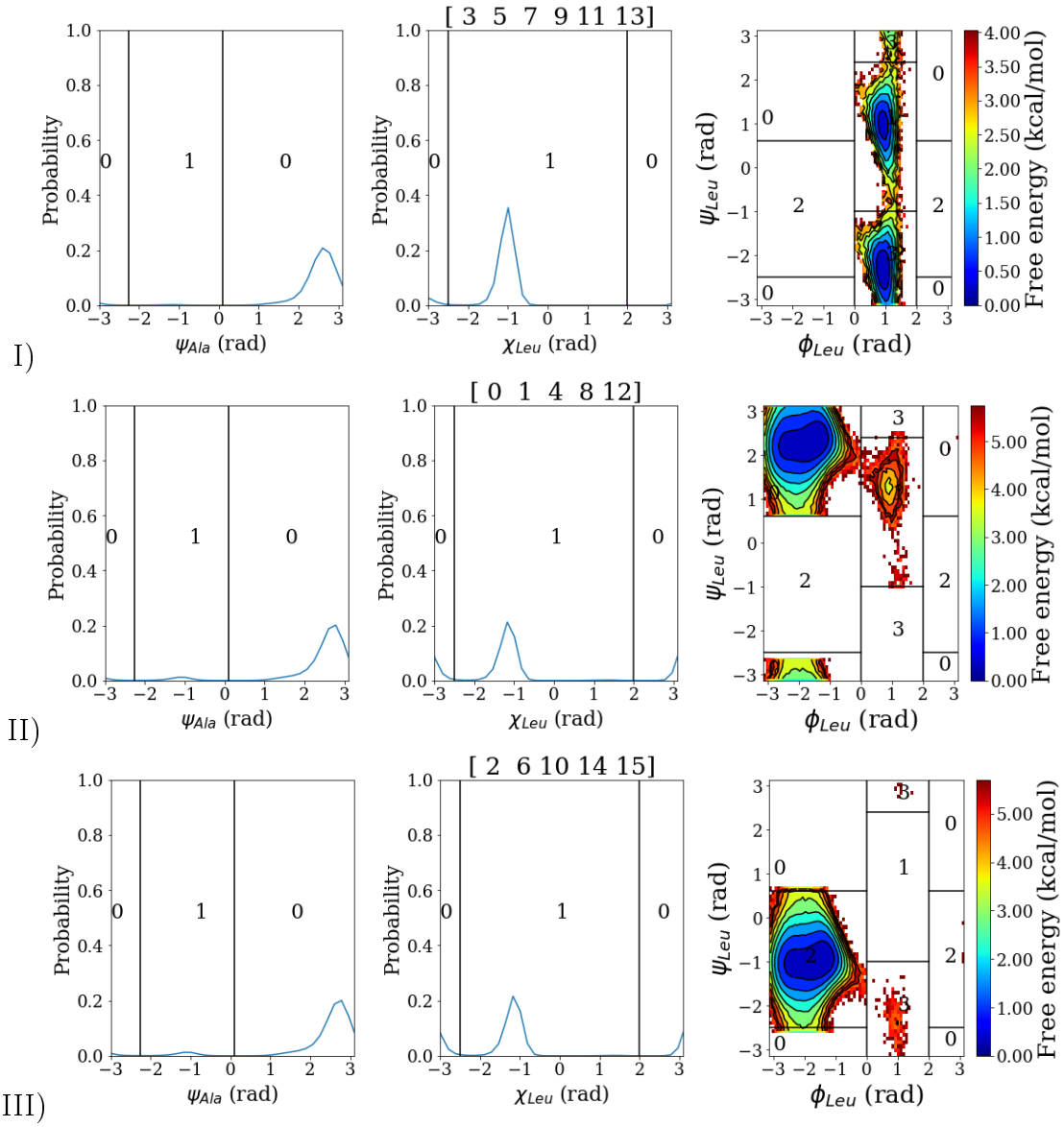


Figure S2: Distribution of torsion angles ψ_{Ala}, χ_{Leu} , ϕ_{Leu}, and ψ_{Leu} in the conformational clusters of the meta-stable sets I,II, and III, as obtained from the classical MD simulation of Ala-Leu in water.

Table S2: Torsion angles of the representative conformations of the most probable micro-states from the classical MD simulations, shown in Figure 3 and subjected to further first-principles simulations (except for cluster 5).

Micro-state	ψ_{Ala}(rad)	χ_{Leu} (rad)	ϕ_{Leu} (rad)	ψ_{Leu}(rad)
0	2.48634	-3.00789	-1.26109	2.07314
4	2.39830	-1.24703	-1.54770	2.06933
2	2.31969	-3.09427	-1.35842	-1.17242
6	2.59943	-1.25551	-1.53055	-1.17207
5	2.21018	-1.12151	1.03061	1.40457

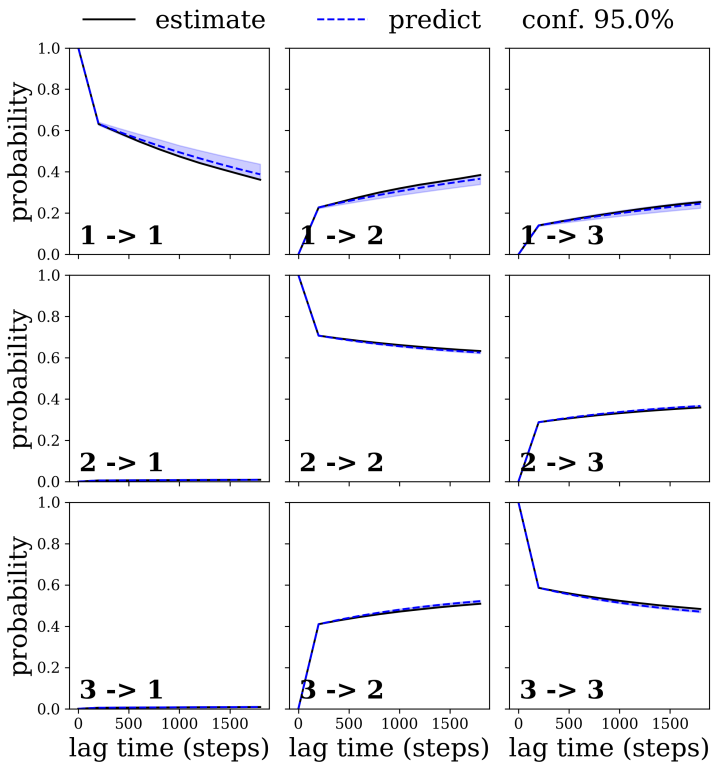


Figure S3: Chapman-Kolmogorov test of the Markov state model (MSM) with three states 1,2,3, corresponding to meta-stable sets I,II,III, respectively. Using the MSM estimated at lag time τ, transition probabilities are predicted for $n \cdot \tau$, and compared to an estimate of the model at lag time $n\tau$.

Normal mode calculations

For each cluster of conformations with a significantly high probability, we performed a geometry optimisation (convergence criterion 3.00E-04 $E_H/\text{Å}$) and subsequent normal mode calculation in implicit water (modelled by a polarisable continuum model, PCM, with a dielectric of ε=80) at the DFT level of theory. The BLYP exchange-correlation density functional and a $6-31G(d)$ basis set was employed for all these static calculations using the Gaussian programme package[8]. The N-terminus and the amino group of the peptide was set deuterated (ND3 and ND). All normal modes were scaled by 0.992 in frequency and then shifted by 85 cm-^{1} shift so as to match best the most intense band in the experimental spectrum.

The IR-spectra computed from the normal-modes in implicit water are shown in Figure S4, top (dashed lines), together with the optimised geometries. With regard to the backbone conformation, all cluster snapshots converge to the same state which allows the charged COO and the polar ND group, and the CO and the terminal ND_3 group to come close to each other and interact optimally. The optimised structures of conformations 0 and 2 and 4,6 differ, however, in their leucine side chain conformations. It is interesting to note, that this side chain appears to have an impact on the frequencies of the CO stretch vibrations. The bands assigned to the C=O stretch and the asymmetric COO-stretch vibrations of the Ala-Leu peptide are two separated bands in the conformers 0 and 2, i.e. with the leucine side chain trans to the peptide bond. In the gauche conformations, 4 and 6, the two vibrational bands are much closer to each other.

Normal modes computed for "microsolvated" Ala-Leu, that is with a total of seven water molecules added, one of each at each hydrogen bond donor or acceptor in the polar groups, ND_3 , CO, ND, and COO, respectively are shown as dotted grey lines in Figure S4, top.

The effect of adding one water molecule at a time at each polar group, ND_3, CO, ND,

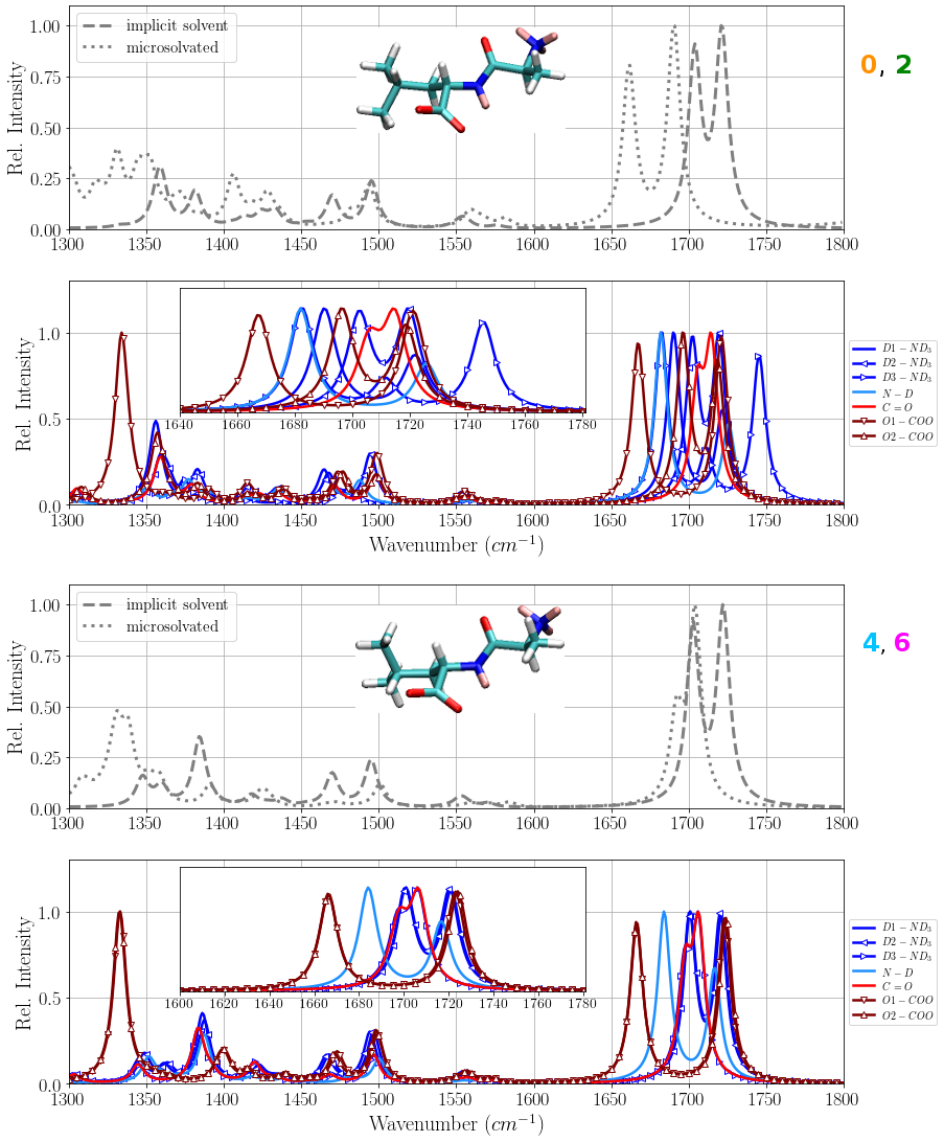


Figure S4: IR spectra computed in harmonic approximation for optimised geometries of the four conformers, corresponding to the most probable micro-states 0, 2, 4, and 6, respectively. Since conformations 0 and 2 and 4 and 6, respectively, are chemically equivalent only one of them is shown, and labelled as 0, 2 and 4, 6,respectively. Top: normal modes of Ala-Leu in implicit solvent (dashed) and additionally hydrogen-bonded to seven water molecules bound to the polar groups, ND3, CO, ND, and COO (dotted). Bottom: Normal modes of Ala-Leu in implicit solvent with one water molecule at a time hydrogen-bond to the polar groups, ND3 (dark blue), CO (red), ND (light blue), and COO (maroon), respectively.

and COO, respectively, is shown in the bottom traces of Figure S4. Addition of a water molecule to the C=O group of conformers 0,2, leads to a red-shift of the corresponding stretch vibration. Addition of a water molecule to the C=O group in conformations 4,6 results in a slightly more pronounced red-shift of the CO band. Addition of a water molecule to the carboxyl group has also a similar effect on all conformers, i.e. the appearance of a rather strong COO band at $\sim$ 1370 cm^{-1}. For the frequencies of the N-D stretch vibrations in all conformations upon addition of a water molecule to the respective group another large red-shift, bringing this band to the amide I region, is observed .

Conformational analyses

Time series of torsion angles and conformational cluster states in the first-principles MD simulations

Table S4: Parts of first-principles trajectories of conformational clusters 0, 2, 4, and 6, respectively, used for the computation of spectra. The labels of the runs refer to the conformation in which the simulation was initiated.

0	2	4	6
0-run2-(0-20ps)	2-run1-(3.5-19ps)	0-run1-(34-56ps)	2-run3-(2-36ps)
0-run4-(0-20ps)	2-run1-(21-30ps)	4-run1-(30-50ps)	4-run1-(60-82ps)
2-run1-(10-50ps)	2-run2-(4-30ps)	6-run4-(0-24ps)	4-run3-(26-52ps)
2-run3-(0-20ps)			4-run4-(26-53ps)
2-run4-(0-20ps)			
6-run1-(10-50ps)			

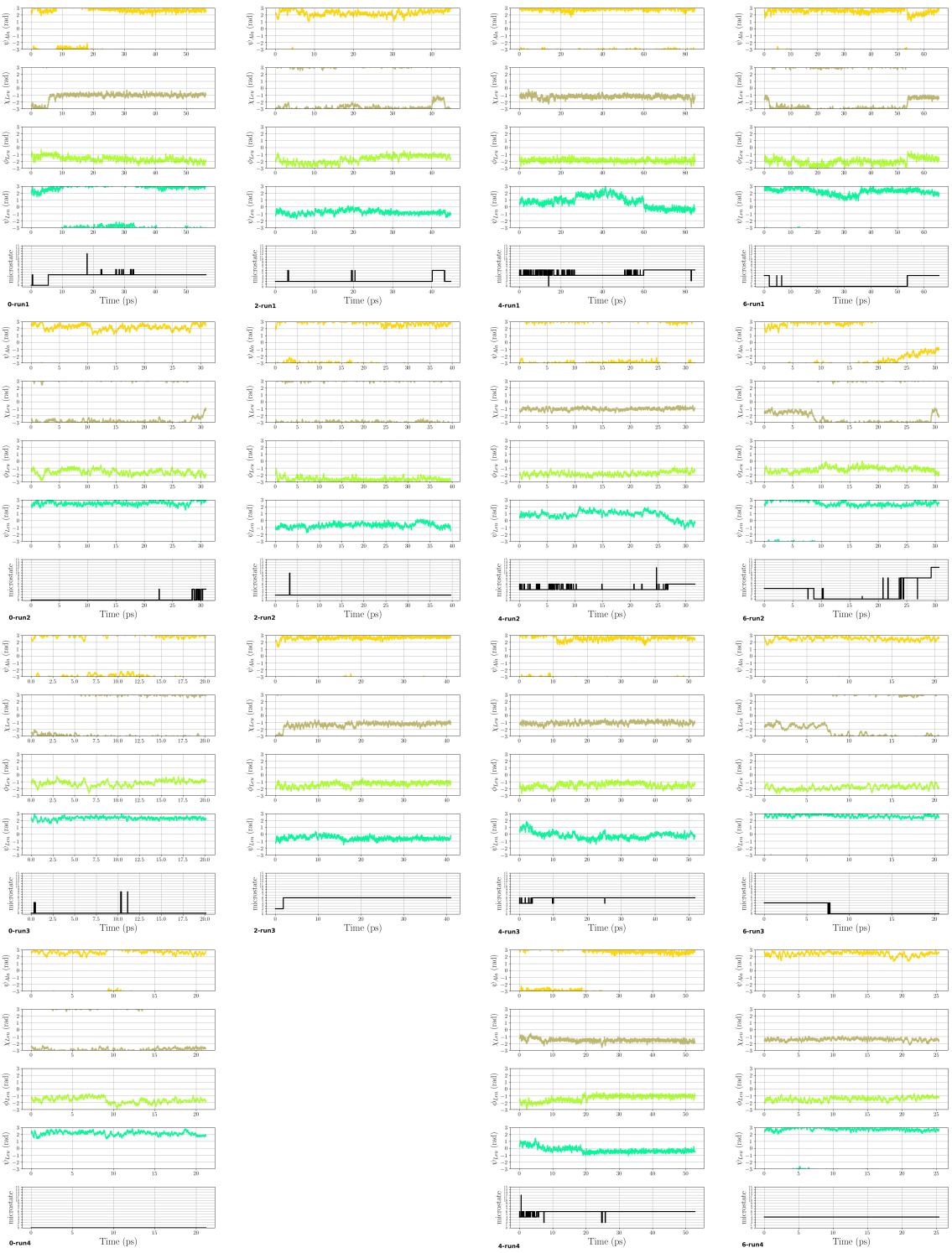
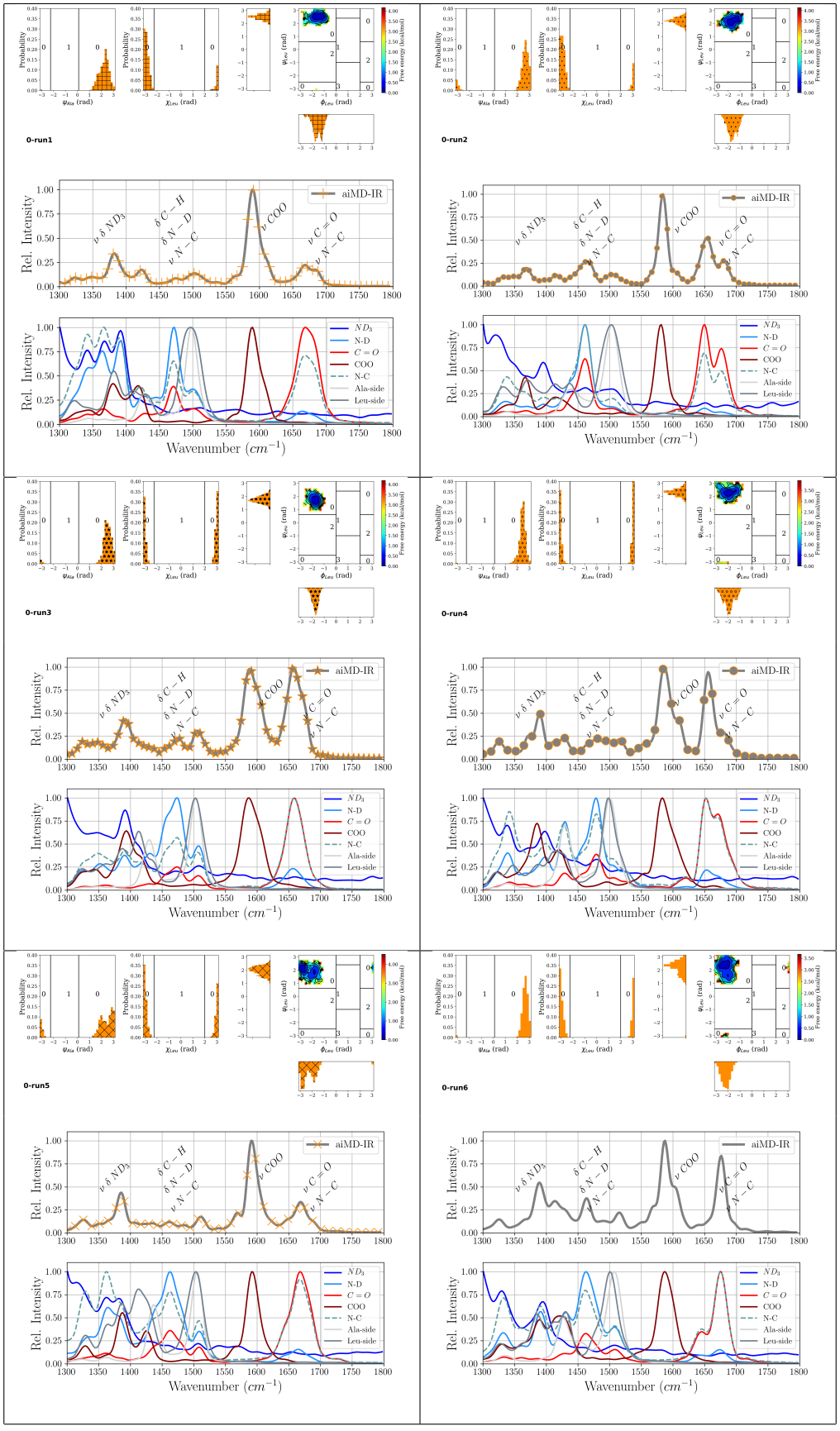


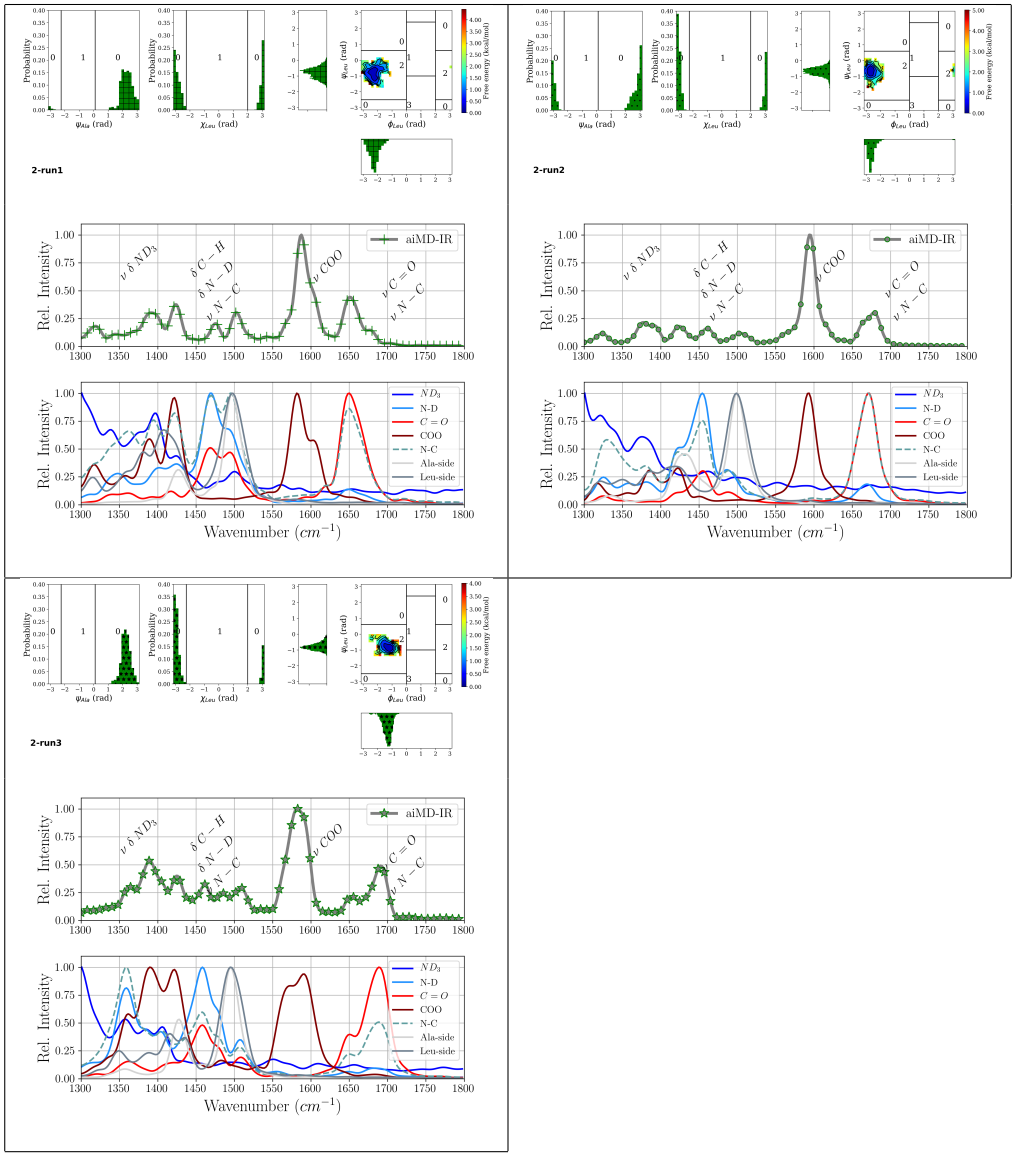
Figure S5: Time series of torsion angles ψ_{Ala}, χ_{Leu} , ϕ_{Leu} and ψ_{Leu} , and conformational states defined by them along the first-principles simulations. Note that a jump between a conformational state is defined by the cluster it belongs to based on the initial discretisation. Such a jump may in fact be only a small change in one torsion angle.

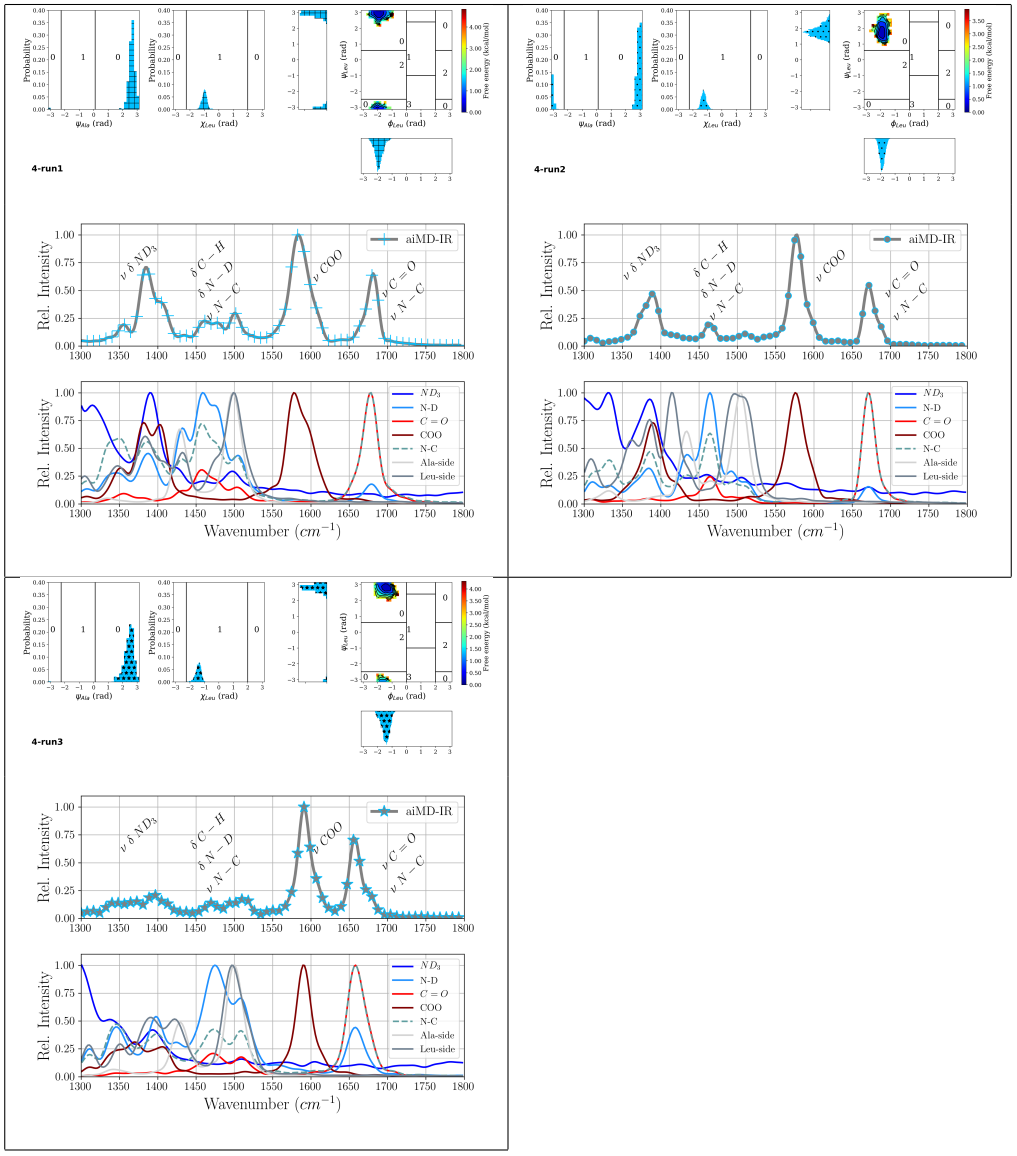
Hydrogen bond distributions

Figure S6: Probability distribution of number of hydrogen-bonds of the peptide with water, analysed separately for ND3, CO, ND, and the two carboxyl oxygen atoms (COO1 and COO2, respectively). The four micro-states 0, 2, 4, and 6 are shown in orange, green, blue, and magenta, respectively. Each row of a sub-figure represents an individual simulation of that micro-state.

Torsion angle distributions, IR and power spectra from first principles MD simulations


0-run1
0-run2
0-run3
0-run4
0-run5
0-run6
aiMD-IR
Rel. Intensity
Wavenumber (cm^{-1})
$\nu, \delta\, ND_3$
$\delta\, C-H$
$\delta\, N-D$
$\nu\, N-C$
$\nu\, COO$
$\nu\, C=O$
$\nu\, N-C$
ND_3
N-D
$C=O$
COO
N-C
Ala-side
Leu-side


2-run1
2-run2
2-run3
aiMD-IR
Rel. Intensity
Wavenumber (cm^{-1})
ν δ ND_3
δ $C-H$
δ $N-D$
ν $N-C$
ν COO
ν $C=O$
ν $N-C$
ND_3
N-D
$C=O$
COO
N-C
Ala-side
Leu-side


4-run1
4-run2
4-run3
aiMD-IR
Rel. Intensity
Wavenumber (cm^{-1})
$\nu, \delta ND_3$
$\delta C-H$
$\delta N-D$
$\nu, N-C$
ν COO
$\nu, C=O$
$\nu, N-C$
ND_3
N-D
$C=O$
COO
N-C
Ala-side
Leu-side

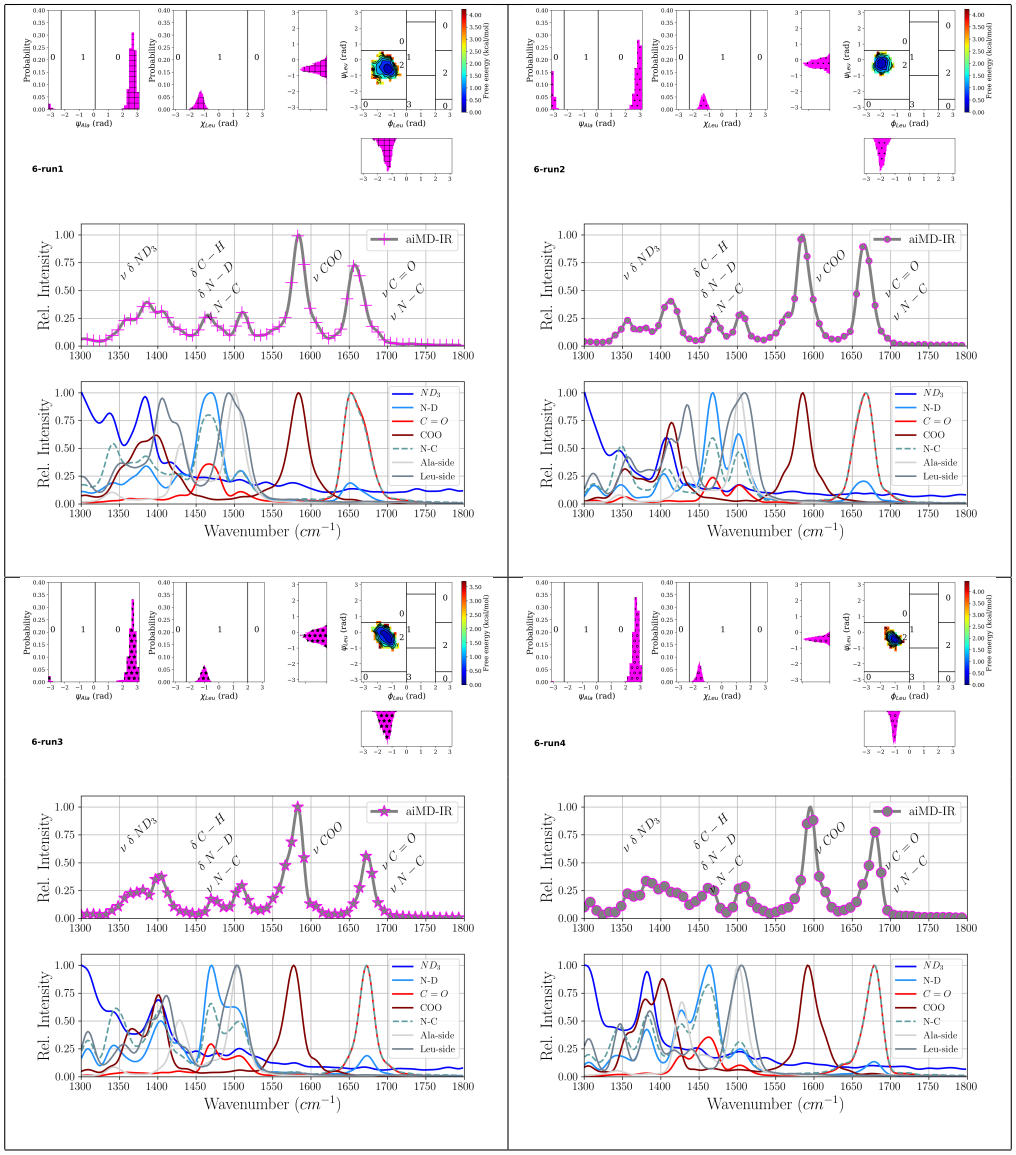


Figure S7: Torsion angle distributions (top), IR spectra (middle) and power spectra (bottom), obtained from the individual first-principles MD simulations of representative Ala-Leu conformations in water. The colours in distribtions and the IR spectra correspond to the different clusters as 0 (orange), 2 (green), 4 (purple), and 6 (magenta), respectively, with all runs of one cluster shown next to each other (run1 ... run3, run4, or run6, respectively, from left to right). The colours in the power spectra correspond to the different groups of atoms. The leucine backbone torsion angles, as shown in the ramachandran plot, would correspond to β-sheet for cluster 0 and 4, and to α-helix for clusters 2 and 6, respectively, but are chemically equivalent conformations due to the second oxygen atom in the carboxyl terminus.

References

(1) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(2) Hornal, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **2006**, *65*, 712–725.

(3) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(4) Darden, T.; York, D.; Pedersen, L. G. Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

(5) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577–8503.

(6) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101+.

(7) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, btt055.

(8) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.;

Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ã.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.