Supporting Information: Exploring the conformational preferences of 20-residue peptides in isolation: Ac-Ala$_{19}$-Lys + H$^+$ vs. Ac-Lys-Ala$_{19}$ + H$^+$ and the current reach of DFT

Franziska Schubert[1], Mariana Rossi[1,2], Carsten Baldauf[1], Peter Kupser[1], Kevin Pagel[1], Frank Filsinger[1], Mario Salwiczek[3],* Beate Koksch[3], Gert von Helden[1], Gerard Meijer[1,4], Matthias Scheffler[1], and Volker Blum[1,5]

[1] *Fritz-Haber-Institut der Max-Planck-Gesellschaft, D-14195 Berlin, Germany*
[2] *Physical and Theoretical Chemistry Laboratory, University of Oxford, OX1 3QZ Oxford, UK*
[3] *Institut für Chemie und Biochemie - Organische Chemie, Freie Universität Berlin, D-14195 Berlin, Germany*
[4] *Radboud University Nijmegen, 65000 HC Nijmegen, The Netherlands and*
[5] *Department of Mechanical Engineering and Materials Science and Center for Materials Genomics, Duke University, Durham, NC 27708, USA*

(Dated: February 15, 2015)

1. EXPERIMENTS

Peptide Synthesis and Purification

All peptides were synthesized automatically with a SyroXP-I peptide synthesizer (MultiSynTech GmbH, Witten, Germany) on a 0.05 mmol scale applying standard Fmoc/tBu-chemistry [1] using Fmoc-Ala-preloaded NovaSyn®-TGA resin (loading 0.22 mmol/g, NovaBiochem). Fmoc-Ala-OH*H_2O and Fmoc-Lys(Boc)-OH were purchased from Orpegen Pharma. All couplings were performed using 4 eq. of the respective Fmoc-amino acid (0.5 M solutions containing 0.8 M $NaClO_4$, 98 %, extra pure, Acros Organics) in DMF (for analysis, 99.5 %, Acros Organics), equal molar amounts (according to the amount of amino acid) of 1-hydroxybenzotriazole (HOBt, Fa. Franz Gerhard & Co) and O-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium tetrafluoroborate (0.33 M in DMF) and a two-fold excess of *N,N*-diisopropylethyl amine (98 %, Acros Organics, 2 M in NMP, 99.5 %, for peptide synthesis, Acros Organics). Couplings were performed as double-couplings, 30 minutes each and Fmoc-deprotection was achieved by treating the resin four times (five min. each) with 4 mL of a DMF solution of 2 % (v/v) piperidine (extra pure, 99 %, Acros Organics) and 2 % (v/v) 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU, Merck). Capping was performed using 10 % (v/v) freshly distilled acetic anhydride (99 %, Acros Organics) and 10 % (v/v) N-ethyldiisopropylamine (DIEA, 98+ %, Acros Organics) for 10 minutes (three times). For the formylated peptides mentioned below, formylation was achieved by treating the resin over night with 3 mL of ethylformiat (Merck)[2]. Resin-cleavage and side-chain deprotection were achieved by treating the resins 3-4 hours with 4 mL of a mixture of 95 % (v/v) trifluoroacetic acid (TFA for synthesis, Merck), 2.5 % (v/v) deionized water (Millipore®), and 2.5 % (v/v) triisopropyl silane (TIS, 99 %, Acros Organics). The resin was washed twice with 2 mL TFA and the excess TFA was evaporated by a stream of nitrogen. The oily residue was suspended in 30-50 mL deionized water and extracted three times with 30-50 mL diethylether (Et_2O, prolabo, VWR). After removal of residual ether *in vacuo* the peptides were lyophilized and used without further purification.

Infrared Photodissociation Experiments

The experiments were performed at the free electron laser facility FELIX [3] (Nieuwegein, the Netherlands) using the Fourier-transform ion cyclotron (FT-ICR) mass spectrometer [4]. Prior to analysis the lyophilized peptides where dissolved in 25 % H_2O / 75 % TFA and subsequently diluted to the desired concentration (typically 250 μM). The solution was sprayed via a syringe pump Harvard Apparatus 11plus (Harvard Apparatus, Holliston, MA, USA) and a standard electrospray ionization (ESI) source (Waters Corporation, Milford, MA, USA) connected to the mass spectrometer. The ESI generated ions were transported and accumulated in a hexapole ion trap and subsequently transferred into a home-built FT-ICR mass spectrometer that is optically accessible via a KRS-5 window at the back end. After trapping and mass-selective isolation of the charged molecules of interest inside the ICR cell, the ions were irradiated by IR photons of the free electron laser FELIX. When the IR light is resonant with an IR active vibrational mode in the molecule, this results in the absorbance of many photons, which causes dissociation of the ion (IRMPD). Monitoring the fragmentation yield of the parent ion signal as a function of IR wavelength leads to the IR spectra. Recent studies showed that the so-obtained spectra are not entirely identical to linear IR absorption spectra, but can be very close to them [5]. The output of FELIX is continuously tunable over a range of 40 to 2000 cm^{-1}. In the presented experiment only the range from 1000 to 1800 cm^{-1} was scanned. The light consists of macropulses of about 5 μs length at a repetition rate of 10 Hz, which contain 0.3-

5 ps long micropulses with a micropulse spacing of 1 ns. In the present experiment, macropulse energies were in the range of 35 mJ with a bandwidth of approx. 1 %.

Test for existence of doubly-charged Ac-Lys-Ala$_{19}$ + H$^+$ dimers in IR spectroscopy

In order to test for the existence of doubly charged dimers in the experimental beam, we utilized a mixture containing equal amounts of Ac-Lys-Ala$_{19}$ + H$^+$ and For-Lys-Ala$_{19}$ + H$^+$. In the present work, the For-Lys-Ala$_{19}$ + H$^+$ peptides were *only* employed for the particular test described here, not in any other context. Both peptides exhibit an identical sequence and merely differ in the nature of the N-terminal capping, resulting in a mass-difference of 14 Da. In case of a significant population of dimeric species, mixed hetero oligomers are expected to give rise to a signal at 1531.8 Da. Isolated monomers on the other hand exhibit masses of 1524.8 Da for the formylated and 1538.8 Da for the acetylated species, respectively. Electrospraying the mixture at moderate concentration and the same experimental conditions used to record IRMPD spectra, exclusively showed isolated monomers. Interestingly, the same behavior was observed at very high concentrations (up to 500 μM peptide) which are explicitly known to facilitate the formation of non-specific peptide aggregates [6]. Consequently we conclude that, at least for the instrumental setup utilized here, dimers where not populated to a measureable extent during the experiment.

Ion Mobility-Mass Spectrometry

Ion mobility mass spectrometry (IM-MS) experiments using a Synapt G2-S travelling-wave IM-MS instrument were performed in order to analyze the multiple conformers of Ac-Lys-Ala$_{19}$ + H$^+$ on a semi quantitative scale (clear intensities but no simple quantitative conversion of measured drift times into CCS). This data is shown in Fig. 1 of the main paper. Samples were dissolved in 25% H_2O / 75% TFA to yield a concentration of approximately 50 μM. Subsequently the solution was transferred into in-house prepared quartz glass capillaries and ionized using a nano electrospray ionization (nESI) source. Typical settings of the Synapt G2-S instrument were: source temperature 20 °C; capillary voltage, 0.8-1.2 kV; sample cone 50-100 V; source offset, 20 V; cone gas, off; trap collision energy, 2 V; transfer collision energy, 4 V; trap gas flow, 2 mL/min; helium cell gas flow, 180 mL/min; IM-MS gas flow, 90 mL/min; trap DC bias 45 V; IM-MS wave height, 40 V; IM-MS wave velocity, 500-900 m/s.

We carefully adjusted the "cone" and "source offset" voltages prior to measurements in order to avoid unintended collisional activation of the ions at the front end of the instrument. At the utilized conditions (50-100 V), we did not observe any evidence for collision induced unfolding. Although these values may seem relatively high for a 20-residue peptide, it is important to mention that the G2S instrument is equipped with a step wave ion guide, which does not have the conventional cone-skimmer design used in previous Synapt generations. Also using other molecules we found that in-source activation with this setup generally requires considerably higher voltages for "cone" and "source offset".

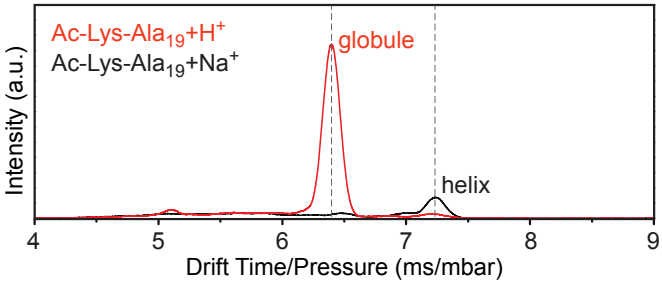

FIG. S1. Arrival time distributions for protonated (red) and sodiated (black) Ac-Lys-Ala$_{19}$-OH using the in-house built drift tube (DT) IM-MS apparatus. Drift times were divided by the pressure inside the DT to account for slight variations in pressure between measurements. These data are consistent with those obtained at the commercial Synapt instrument (see Figure 1 in main paper).

For IM-MS measurements, the source block was not heated and kept at room temperature. At the FT ICR, the source temperature was slightly increased to 60 °C to aid desolvation. However, once the ions enter the instrument they rapidly thermalize back to room temprature via collisional cooling.

To determine absolute collision cross sections (CCSs), we used an IM-MS apparatus constructed in-house following the design described in Ref. 7. Generally, the data from this instrument (e.g., the data shown in S1) show similar results as those obtained at the Synapt instrument (Fig. 1 in paper). CCS values for the corresponding globular and helical monomer forms were calculated from the IM-MS measured drift times as described previously [8]. These CCSs are given in Table III of the main paper.

2. SIMULATIONS

Structure relaxations with DFT (PBE+vdW^{TS})

Structure relaxations with the PBE+vdW^{TS} functional were performed in two steps. First, the optimizations were carried out with light computational settings and then the lowest-energy structures were further relaxed with tight computational settings. As illustrated in Fig. S2, the energy hierarchies obtained with PBE+vdW^{TS} and light settings and with tight settings

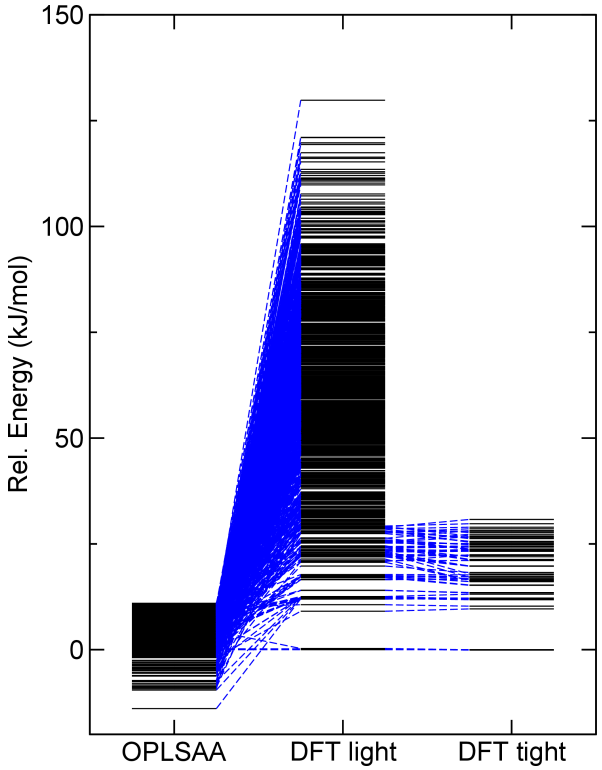

FIG. S2. Ac-Lys-Ala$_{19}$ + H$^+$: Energy hierarchies (black horizontal lines) of OPLSAA conformers and the corresponding structures relaxed with PBE+vdW^{TS} and light computational settings and tight computational settings, respectively. All energies are given relative to the conformer with the lowest energy according to PBE+vdW^{TS}.

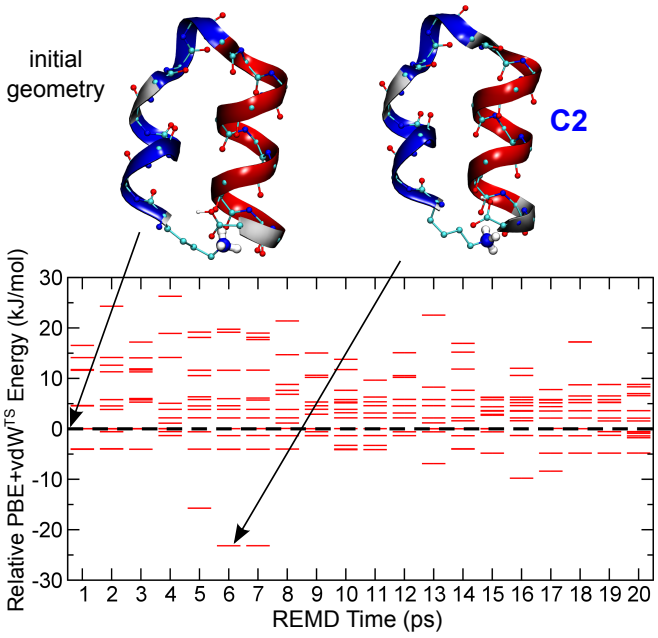

FIG. S3. Example of a particular *ab initio* replica exchange run where lower energy hydrogen bond networks than the "initial geometry" are found. After each ps, a snapshot of each replica is fully relaxed with PBE+vdW^{TS}. Red bars: Energy of each replica snapshot after full PBE+vdW^{TS} relaxation, relative to the energy of the initial structure.

change only very little. However, the hierarchies for the OPLSAA force field and the PBE+vdW^{TS} (light settings) change significantly, which has also been found for other peptide systems before[9–11].

First-principles REMD

Ac-Lys-Ala$_{19}$ + H$^+$: We performed PBE+vdW^{TS} REMD runs for the 4 lowest energy (PBE+vdW^{TS}) structures. For this we used 16 replicas with temperatures ranging between 300 K and 623 K according to a geometric distribution and employed the Nosé-Hoover thermostat [12, 13] using a swapping attempt frequency of 100 fs and a time step of 1 fs (total simulation time: 4x320 ps). After each ps of REMD time we relaxed the snapshots of all 16 replicas with PBE+vdW^{TS}. Fig. S3 shows the energies of all relaxed replicas for a particular example of an *ab initio* REMD run for Ac-Lys-Ala$_{19}$ + H$^+$ where a local rearrangement occured. The lowest energy conformation obtained is more than 20 kJ/mol lower in energy than the starting conformation. The latter structure and the initial conformation are overall very similar with only subtle differences close to the terminations.

Ac-Ala$_{19}$-Lys + H$^+$: For Ac-Ala$_{19}$-Lys + H$^+$ we performed a PBE+vdW^{TS} REMD run for the lowest-energy structure. We used again 16 replicas with temperatures ranging between 300 K and 623 K according to a geometric distribution. Different from the simulation for Ac-Lys-Ala$_{19}$ + H$^+$, we here employed the Bussi-Donadio-Parrinello thermostat [14] with a swapping attempt frequency of 200 fs (total simulation time: 1x208 ps). Those differences in the settings do not have a specific reason, but occurred because the simulations were performed at different times. However, it was too expensive to recompute the REMD run for Ac-Ala$_{19}$-Lys + H$^+$ just for consistency, especially as we do not expect any impact on the results. In fact, for Ac-Ala$_{19}$-Lys + H$^+$ no conformer that was lower in energy than the initial structure was found during the *ab initio* REMD run. Please keep in mind that the global sampling has been performed in the previous step based on force fields, and the *ab initio* REMD runs are used for a local sampling.

Role of a better force field: AmoebaPro13

We relaxed the Ac-Lys-Ala$_{19}$ + H$^+$ conformers C1 to C6 with the AmoebaPro13[15] force field. For this, we used version 6.2 of the TINKER program[16]. As shown in Fig. S4, the prediction of AmoebaPro13 sig-

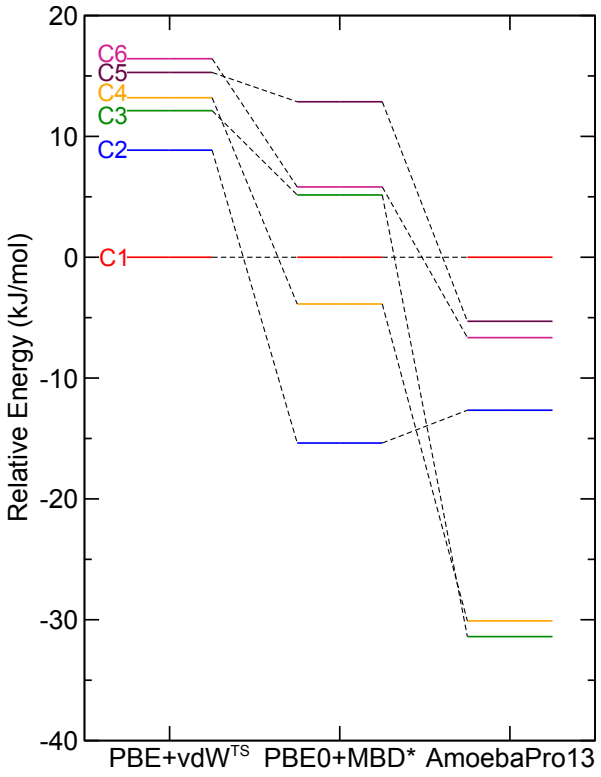

FIG. S4. Energy hierarchies (horizontal bars) for the Ac-Lys-Ala$_{19}$ + H$^+$ monomers C1 to C6 obtained with the PBE+vdW$^{\mathrm{TS}}$ and the PBE0+MBD* functionals in comparison to the AmoebaPro13 force field. The energies are given relative to C1. The dashed lines serve as a guide to the eye.

nificantly differs from the results of PBE+vdW$^{\mathrm{TS}}$ and PBE0+MBD*. For AmoebaPro13, the C3 conformer has the lowest energy, while C4 follows very closely. The C1 conformer, which is the lowest one in PBE+vdW$^{\mathrm{TS}}$, has the highest energy in AmoebaPro13.

Ac-Lys-Ala$_{19}$ + H$^+$: helical models

We also performed pure first-principles searches to find typical conformations of α-helical model peptides Ac-Lys-Ala$_{19}$ + H$^+$. The proton can be either located close to the C-terminus or at the N-terminal lysine. For the latter type we performed 8 ps of *ab initio* REMD using 16 replicas (total simulation time: 128 ps) with temperatures ranging between 300 K and 623 K according to a geometric distribution, starting from a perfect α-helical conformation. For the case with the proton located close to the C-terminus, we performed 32 ps *ab initio* REMD using 18 replicas (total simulation time: 576 ps) with temperatures ranging between 300 K and 688 K, starting from three different conformations where the proton is associated with the backbone carbonyl oxygen atom no. 17, 18, and 19 (counted from the N-terminus), respectively. The lowest energy conformers for both the proton located at the N-terminal lysine and close to the C-terminus are depicted in Fig. 4 of the main manuscript together with their (free) energies relative to the lowest energy monomer conformer C1.

Relaxations with PBE0+vdW$^{\mathrm{TS}}$ and PBE0+MBD*

For reasons of computational feasibility, relaxations with the PBE0-based functionals were not performed using tight settings. Instead, we employed LVL-intermediate settings and then followed up with single-point calculations using tight settings. The LVL-intermediate settings inherit the grid definitions of tight settings but only contain the first basis function of tier2 in addition to tier1. Additional basis functions are used for the computation of exchange integrals. Conventional tight settings comprise the full tier2 basis sets[17].

Comparison of infrared spectra and the Pendry R-factor

In the main paper, we use the Pendry R-factor to quantitatively compare measured IRMPD spectra and computed infrared spectra based on *ab initio* molecular dynamics and the dipole-dipole approximation (Figure 4, main paper).

We have two independent measurements for the IR spectrum of Ac-Lys-Ala$_{19}$ + H$^+$, exp. (1) and exp. (2). Both are very similar (see Fig. S5) although they originate from completely different measurement cycles. Exp. (1) is the one shown in the main paper.

As the Pendry R-factor is sensitive to small kinks in the spectra, the experimental data had to be smoothed before comparing to the calculated spectra. In order not to oversmooth the data we first splined the experimental raw data on a grid with a width of 2 cm^{-1}. Afterwards, the data was smoothed twice (3-point-formula) and interpolated onto a fine 0.5 cm^{-1} grid. Figure S5 depicts the raw data for both experiments (1) and (2) for Ac-Lys-Ala$_{19}$ + H$^+$ versus the processed data.

R_{P} is calculated including a rigid shift Δ_x (along the wavenumber axis) and Δ_y (along the normalized intensity axis) of the theoretical spectrum that give the best agreement with experiment. Δ_x most likely reflects a systematic mode softening due to the chosen exchange correlation functional and missing nuclear quantum effects. As mentioned in the main paper, the absorption of several photons in IRMPD experiments leads to a broadening of the peaks and can affect relative band intensities. Thus the observed intensity differences between spectrum (1) and (2) can be attributed to different tunings of the laser since relative band intensities can be sensitive to the precise experimental conditions. Δ_y accounts for this. Thus, including the shifts Δ_x and Δ_y

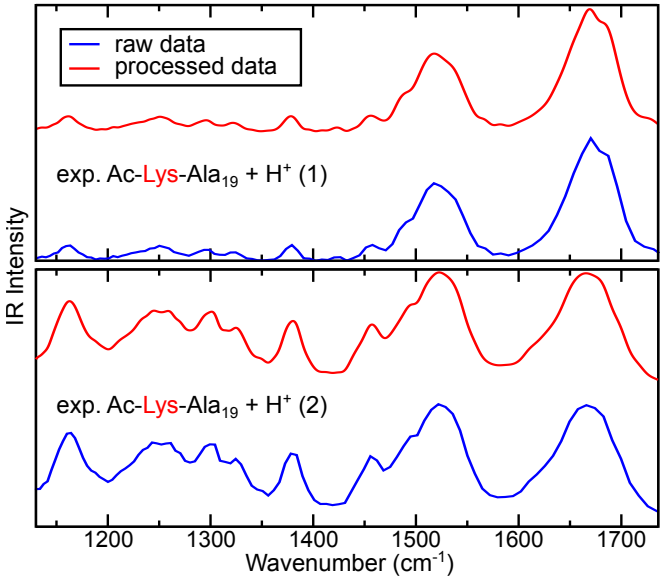

FIG. S5. Raw versus processed experimental data for both IR measurements of Ac-Lys-Ala_{19} + H^+. The spectra are normalized to the highest peak.

reflects not a "fitting" of parameters but rather allows us to eliminate small, possible systematic uncertainties, that are hard or impossible to quantify. As a result, the reported R_P focus on what is systematically important in the spectra. All details are given in Tab. I.

* Present address: CSIRO Materials Science and Engineering, Bayview Avenue, Clayton, Victoria 3168, Australia

[1] G. B. Fields and R. L. Noble, Int. J. Pept. Protein Res. **35**, 161 (1990).
[2] S. T. Breitung, J. J. Lopez, G. Dürner, C. Glaubitz, M. W. Göbel, and M. Suhartono, Beilstein J. Org. Chem. **4**, 35 (2008).
[3] D. Oepts, A. van der Meer, and P. van Amersfoort, Infrared Phys. Technol. **36**, 297 (1995).
[4] J. J. Valle, J. R. Eyler, J. Oomens, D. T. Moore, A. F. G. van der Meer, G. von Helden, G. Meijer, C. L. Hendrickson, A. G. Marshall, and G. T. Blakney, Rev. Sci. Instrum. **76**, 023103 (2005).
[5] J. Oomens, B. G. Sartakov, G. Meijer, and G. von Helden, Int. J. Mass Spectrom. **254**, 1 (2006).
[6] J. L. P. Benesch, B. T. Ruotolo, D. A. Simmons, and C. V. Robinson, Chem. Rev. **107**, 3544 (2007).
[7] P. R. Kemper, N. F. Dupuis, and M. T. Bowers, Int. J. Mass Spectrom. **287**, 46 (2009).
[8] E. Mason and W. McDaniel, *Transport properties of ions in gases* (Wiley, 1988).
[9] M. Rossi, S. Chutia, M. Scheffler, and V. Blum, J. Phys. Chem. A **118**, 7349 (2014).
[10] C. Baldauf, K. Pagel, S. Warnke, G. von Helden, B. Koksch, V. Blum, and M. Scheffler, Chem. Eur. J. **19**, 11224 (2013).
[11] M. Rossi, M. Scheffler, and V. Blum, J. Phys. Chem. B **117**, 5574 (2013).
[12] S. Nosé, J. Chem. Phys. **81**, 511 (1984).
[13] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).
[14] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 014101 (2007).
[15] Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, J. Chem. Theory Comput. **9**, 4046 (2013).
[16] J. W. Ponder, "Tinker - software tools for molecular design," http://dasher.wustl.edu/ffe/.
[17] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, and M. Scheffler, Comput. Phys. Commun. **180**, 2175 (2009).

TABLE I. R-factors for the spectra of the conformers C1 to C4, and the Ac-Lys-Ala_{19} + H^+ helix with the proton located close to the C-terminus against both experimental spectra for the wavenumber range between 1130 and 1736 cm^{-1}. Additionally, the rigid shifts along the wavenumber axis (Δ_x) and the (normalized) intensity axis (Δ_y) are listed.

	experiment 1			experiment 2		
conformation	R_P	Δ_x (cm^{-1})	Δ_y	R_P	Δ_x (cm^{-1})	Δ_y
C1	0.44	25.0	0.000	0.31	25.0	0.025
C2	0.31	21.5	0.000	0.23	22.5	0.025
C3	0.33	22.0	0.000	0.33	24.5	0.050
C4	0.34	24.5	0.000	0.24	25.0	0.035
Ac-Lys-Ala_{19} + H^+, helix, (H^+ near C-term.)	0.29	19.5	0.000	0.27	20.0	0.020