

Structure of Sodiated Octa-Glycine: IRMPD Spectroscopy and Molecular Modeling

David Semrouni,^a O. Petru Balaj,^{a,*} Florent Calvo,^b Catarina F. Correia,^{a,†} Carine Clavaguéra,^a and Gilles Ohanessian^a

^a Laboratoire des Mécanismes Réactionnels, Ecole Polytechnique and CNRS, Palaiseau, France

^b Laboratoire de Spectrométrie Ionique et Moléculaire, Université de Lyon I and CNRS, Villeurbanne, France

The structure of the sodiated peptide GGGGGGGG-Na⁺ or G₈-Na⁺ was investigated by infrared multiple photon dissociation (IRMPD) spectroscopy and a combination of theoretical methods. IRMPD was carried out in both the fingerprint and N-H/O-H stretching regions. Modeling used the polarizable force field AMOEBA in conjunction with the replica-exchange molecular dynamics (REMD) method, allowing an efficient exploration of the potential energy surface. Geometries and energetics were further refined at B3LYP-D and MP2 quantum chemical levels. The IRMPD spectra indicate that there is no free C-terminus OH and that several N-Hs are free of hydrogen bonding, while several others are bound, however not very strongly. The structure must then be either of the charge solvation (CS) type with a hydrogen-bound acidic OH, or a salt bridge (SB). Extensive REMD searches generated several low-energy structures of both types. The most stable structures of each type are computed to be very close in energy. The computed energy barrier separating these structures is small enough that G₈-Na⁺ is likely fluxional with easy proton transfer between the two peptide termini. There is, however, good agreement between experiment and computations in the entire spectral range for the CS isomer only, which thus appears to be the most likely structure of G₈-Na⁺ at room temperature. (*J Am Soc Mass Spectrom* 2010, 21, 728–738) © 2010 American Society for Mass Spectrometry

The biological importance of sodium in performing or facilitating essential biological processes, such as neurotransmission, osmotic balance, and cellular metabolism is well documented [1–3]. Mass spectrometric methods have been used extensively to provide insight into peptide sequences [4, 5] starting from sodium-cationized species, however with considerable debate as to the structure of the parent species and the fragmentation mechanisms [6–8]. In this context, sodiated oligoglycines have been used in the last decade as a valuable testing ground for new experimental developments designed to obtain refined energetic and/or structural data. These include ion mobility measurements for global shape information [9, 10], H/D exchange extent and kinetics for isomeric/conformational content [11], the kinetic [12, 13] and the threshold collision induced decomposition [14] methods for thermochemical measurements, as well as infrared multiple photon dissociation (IRMPD) spectroscopy [15] for identification of functional groups and their

interactions. All these studies have been complemented by extensive molecular modeling as required for translating experimental data into properties of specific molecular structures.

Oligoglycines owe their value as model peptides to their relative simplicity. While the number of residues is an obvious source of conformational complexity, the absence of side chains limits the number of factors shaping their structures and energies. On the one hand, the main components of sodium-molecule interactions are electrostatic and polarization, favoring metal binding to the carbonyl oxygens. On the other hand, internal structuration of the peptides is dominated by the construction of a network of hydrogen bonds. The smaller peptides with 2–5 residues tend to wrap around the metal ion to maximize the binding interactions [13]. In larger peptides, a competition takes place between metal-peptide binding and internal stabilization of the peptide through multiple hydrogen bonds [16]. As a result, the most stable structures are much more difficult to predict.

It is the purpose of this paper to report our investigations on one such case, sodium-octaglycine or G₈-Na⁺. We used IRMPD spectroscopy to obtain fingerprints of the functional groups present in G₈-Na⁺, and of their environments. This is carried out in two energy ranges, 1000–1900 and 3000–3700 cm⁻¹. Interpretation of the results required extensive molecular modeling

Address reprint requests to Dr. G. Ohanessian or Dr. C. Clavaguéra, DCMR, Ecole Polytechnique-CNRS, 91128 Palaiseau Cedex, France. E-mail: gilles.ohanessian@polytechnique.fr; carine.clavaguera@dcmr.polytechnique.fr

* Current address: Institut für Physikalische Chemie der Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany/

† Current address: Laboratoire Analyse et Modélisation pour la Biologie et l'Environnement, Université d'Evry Val d'Essonne et CNRS, boulevard F. Mitterrand, 91025 Evry Cedex, France.

using a variety of methods, ranging from extensive sampling with a polarizable force field to quantum chemical calculations using density functional theory and the ab initio MP2 method. We first describe the computational and experimental methods used, then we describe the IRMPD spectra and computed IR absorption spectra, and combine them to ascribe a structure to gaseous $G_8\text{-Na}^+$ at 298 K.

Computational Methods

Replica-Exchange Molecular Dynamics with the AMOEBA Force Field

The salt-bridge and charge-solvated isomers of $G_8\text{-Na}^+$ have been modeled using the polarizable AMOEBA force field [17, 18], which was shown to provide reliable structural and energetic properties for sodiated oligoglycine peptides [16]. In this force field, the electrostatic component incorporates a fixed partial charge, a dipole, and a quadrupole on each atom as derived from quantum mechanical calculations. Many-body polarization effects are explicitly treated using a self-consistent induced dipole polarization procedure. The conformational landscapes were explored using molecular dynamics (MD) simulations, improved with the replica-exchange (RE) strategy. The REMD approach is now widely used as a powerful tool for sampling the potential energy surface of biomolecular systems [19–21]. Briefly, the idea is to perform simultaneously several MD trajectories at various temperatures, occasionally attempting an exchange between two configurations of neighboring trajectories, and accepting this exchange with a Metropolis probability [22]. The success of the REMD method comes from the very broad sampling at high temperatures, where barrier crossing is signifi-

cantly enhanced, combined with an efficient annealing to low-energy structures by successive exchanges with lower temperature replicas. Its popularity is due to the use of conventional simulations in canonical ensembles, rather than nonphysical biases, as well as straightforward parallelization.

The present REMD simulations have been performed with eight replicas allocated according to a geometric progression in the range 200–600 K, that is 200, 234, 274, 320, 375, 438, 513, and 600 K. This progression maximizes the successive overlap between energy distributions (at least for harmonic systems), which is a necessary condition for successfully exchanging configurations. Each molecular dynamics trajectory was propagated with a time step of 1 fs, and the temperature was controlled with a Berendsen thermostat with a coupling time of 0.1 ps. Exchanges of a random pair of adjacent configurations were attempted every 10 ps, and configurations were saved every 4 ps for further local optimization with AMOEBA. The total simulation time for each replica was taken as 4 ns, and the REMD simulations were repeated twice, with all replicas starting each time from the most stable structure found so far. The present simulations with the AMOEBA force field were performed with the Tinker software package [23], recently interfaced for replica-exchange MD simulations by Penev and coworkers [24].

Figure 1 shows the distributions of potential energies obtained for the SB complex during the third REMD iteration. These histograms show a broad overlap for the four highest temperatures, but the four replicas at lowest temperatures seem to overlap poorly. However, looking at the instantaneous index of the replica initially set to the lowest temperature replica (200 K), we find a very satisfactory number of successful exchanges

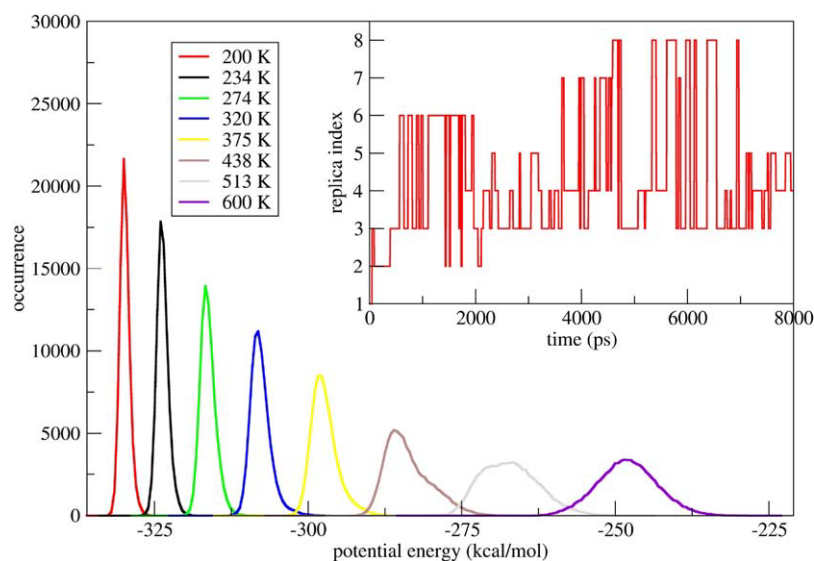


Figure 1. Histograms of potential energy visited during the REMD simulations of the SB system. The inset shows the replica index corresponding to 200 K, indicating that all replicas proceed several times at the lowest temperature within 8 ns.

with other replicas, up to the highest temperature of 600 K. This ensures that our temperature schedule will allow a correct communication between replicas, hence a much improved sampling with respect to conventional MD.

Three iterations of REMD simulations appeared necessary for the SB system, as new minima lower in energy than those previously obtained were found after the first run. The first iteration started from a structure constructed from a previously obtained CS isomer (G_8 _1, see Table 1) by transferring the proton from the C- to the N-terminus. The second iteration was initiated with the lowest energy minimum obtained during the first, while the third iteration was simply a restart from the final configurations of the second iteration. The same computational effort was devoted to the CS system, however, no conformer lower in energy than minima obtained previously by conventional MD simulations [16] could be located.

The most stable configurations collected during the REMD explorations were further subjected to electronic structure calculations to determine the best candidates for comparison to experiment.

Quantum Chemical Calculations

For the conformations selected from REMD simulations, geometry optimizations were performed at DFT-D level with the B3LYP functional and a SVP basis set. The B3LYP functional has been shown to be unre-

liable for relative energies of G_8 -Na⁺ conformers [16] and A_n -H⁺ conformers ($n = 3$ -5,7) [35]. Thus, we used the general empirical dispersion correction proposed by Grimme for density functional calculations [25, 26]. The resolution-of-the-identity (RI) approximation was employed for molecular orbital two-particle integrals. The errors made within this approximation, with optimized auxiliary basis sets, are generally negligible compared with errors due to the one-electron basis set incompleteness. RI-B3LYP-D and RI-MP2 energy calculations on DFT-D optimized geometries were performed with a triple- ζ TZVPP basis set. Vibrational frequencies were calculated at the RI-B3LYP-D/SVP level. The calculated band intensities were convoluted assuming a Lorentzian profile with a 15 cm⁻¹ width. The vibrational frequencies were scaled by a factor 0.96 in the entire spectral range. While it is often considered that different scaling factors should be used below 2000 and above 3000 cm⁻¹, reasonable matching was found to occur with a single scaling factor in the present case. The value of 0.96 is a good compromise for B3LYP-D/SVP frequencies in the amide A region as shown by Bouteiller et al. [27]. Finally, the relative Gibbs free energies at 298 K were determined at the B3LYP-D/TZVPP//B3LYP-D/SVP and MP2/TZVPP//B3LYP-D/SVP levels of theory. In the cases of the lowest salt bridge and charge solvation structures, RI-MP2/SVP geometry optimizations and vibrational frequency calculations were additionally carried out. A scaling factor of 0.943 was applied to MP2 frequencies. All quantum chemical calculations were performed with the TURBOMOLE 5.10 package [28].

Experimental

The ions were generated by electrospray ionization and analyzed using two experimental setups, to record IRMPD spectra in two different photon energy ranges. The 1000–1800 cm⁻¹ fingerprint region generally provides information on the identity of the functional groups in the molecule and to some extent on their environment, while the 2900–3600 cm⁻¹ range is where the O–H and N–H stretching motions are expected to be IR-active with frequencies that are highly sensitive to the existence and strength of hydrogen bonds. In the first case, we used a quadrupole ion trap mass spectrometer (modified Bruker Esquire 3000+) [29, 30] coupled to the Free Electron Laser (FEL) at the Centre Laser Infrarouge d'Orsay (CLIO). In the second case, we used an FT-ICR mass spectrometer (Bruker APEX IV) equipped with a 7 T superconducting magnet, coupled to an OPO/OPA laser also at CLIO [31, 32].

Operations of the FEL and OPO/OPA lasers, coupled to either of the above mentioned mass spectrometers, have been detailed elsewhere [29–32]. We provide below the specific parameters used to obtain the results described herein.

Table 1. Relative energies (kcal/mol) of the lower-energy G_8 -Na⁺ structures at B3LYP-D and MP2 levels. All results are at B3LYP-D optimized geometries except for CS_1. ΔE_o stands for relative electronic energies and ΔG for relative Gibbs free energies

Name	ΔE_o		ΔG	
	B3LYP-D	MP2	B3LYP-D	MP2
G_8 _compact	15.80	16.14	12.10	12.44
G_8 _1	12.76	11.96	12.06	11.25
G_8 _helix	11.35	8.33	10.64	7.61
G_8 _1N	3.77	4.22	6.85	7.30
G_8 _5	8.38	7.12	5.88	4.63
CS_10	6.56	9.73	10.12	13.29
CS_9	12.95	14.00	12.16	13.21
CS_8	13.46	14.85	10.41	11.79
CS_7	1.17	3.98	3.93	6.74
CS_6	5.03	6.02	5.47	6.46
CS_5	1.68	2.26	4.65	5.23
CS_4	-0.50	1.75	2.89	5.14
CS_3	3.46	4.19	4.14	4.86
CS_2	3.40	3.29	4.24	4.13
CS_1	0.82	1.32 (1.40) ^a	1.46	1.96 (0.49) ^a
SB_5	12.89	12.46	14.85	14.42
SB_4	7.55	8.46	8.34	9.26
SB_3	6.11	6.30	8.53	8.73
SB_2	5.09	4.48	7.39	6.78
SB_1	0.00	0.00	0.00	0.00

^aUsing MP2 geometries and vibrational frequencies.

IRMPD Experiments in the Fingerprint Region

The experimental set-up is based on a modified Bruker Esquire 3000+ ion trap mass spectrometer. To allow the irradiation of the ions with the focused CLIO beam, a 0.7 mm hole was drilled in the ring electrode. The ions generated by an electrospray ion source are transferred to the trap by two octopoles. There, the ions are stabilized by collisions with the He bath gas at a pressure of 10^{-3} mbar. After a rapid thermalization, the ions are confined in a very small volume in the center of the trap. The laser beam was focused in the center of the trap using a 500 mm focal length ZnSe focal lens. The laser beam position was optimized to maximize the fragmentation and allowed for a greater overlap with the ion cloud, resulting in increased fragmentation efficiency than in the ICR cell.

The IR FEL CLIO light is delivered in 8 μs long macropulses fired at a repetition rate of 25 Hz. Each macropulse contains 500 micropulses, each a few picoseconds long. The mean IR power was about 500 mW corresponding to micropulse and macropulse energies of 40 μJ and 20 mJ, respectively. For the experiments described herein, we have used an electron energy of 45 MeV, which allows access to wavelengths in the 1000–1900 cm^{-1} range. In a specific experiment, the 1900–2400 cm^{-1} was searched with a laser power of ca. 600 mW. No absorption band could be detected.

Following a 0.5 s relaxation time, the ions are mass selected and fragmented by macropulses of IR light. Ion fragmentation is observable with only 3 macropulses, but in the experiments presented herein the spectra have been recorded with 10 macropulses. The conditions were tuned for maximum fragmentation efficiency while avoiding saturation.

IRMPD Experiments in the 3000–3600 cm^{-1} Region

The ions are irradiated with the beam generated by a OPO/OPA laser (LaserVision, Bellevue, WA, USA) pumped by a 25 Hz Nd:Yag Splitlight 600 (Innolas Laser GmbH, Krailling, Germany). The laser beam enters the FT-ICR cell along the magnet axis and interacts with the ions trapped into 10^{-10} mbar vacuum. A spherical mirror with 2 m focal distance, placed between the laser and the cell, generates an almost constant diameter of the laser beam in the cell region. Once trapped into the ICR cell, the ions were irradiated for 20 s before signal detection. Each point in the spectra was measured only once, without any signal averaging. For maximum fragmentation efficiency, the IRMPD spectra were split in sections no larger than 300–400 cm^{-1} , and the laser power was optimized before each measurement for the range we intended to scan.

IRMPD fragmentation spectra are plotted using the fragmentation yield R defined as: $R = -\ln[\sum I_{\text{Fragment}} / (I_{\text{Parent}} + \sum I_{\text{Fragment}})]$. The FEL fingerprint spectrum was obtained by including the signal intensities for 24

fragment ions, while the OPO/OPA spectrum was based on six fragment ion intensities.

The spectra in both photon energy ranges were smoothed using a weighted polynomial regression smoothing algorithm.

Sample Preparation

Gaseous ions were obtained by electrospray of diluted sample solutions. First, 15 mL peptide stock solutions were prepared using 6 mL 10% formic acid and 9 mL $\text{H}_2\text{O}/\text{MeOH}$ 50:50 to dissolve the 10 mg of solid peptide; 10–50 μL of stock solution (10^{-3} M) were mixed 1:1 with NaCl solution (10^{-3} M) and diluted to 1.0 mL with $\text{H}_2\text{O}/\text{MeOH}$ 50:50 with 1% formic acid. This protocol produced samples with concentrations of peptide in the μM range, leading to the formation of the sodiated peptide ion as the dominant species in the gas-phase.

Results and Discussion

The IRMPD spectra of $G_8\text{-Na}^+$ in the two spectral ranges investigated are shown in Figure 2 and Figure 4, together with those computed for some of the lowest energy structures shown in Figure 3. Band assignment can be made straightforwardly in some cases, for instance the amide I and amide II massifs in the 1650–1750 and 1480–560 cm^{-1} ranges, respectively. The absence of a free carboxylic acid O–H stretching band near 3550 cm^{-1} [33, 34] is equally noteworthy. Yet other band assignments and the relation of these experimental patterns to one or several molecular structures require modeling results.

Previous molecular dynamics searches [16] were carried out only from a CS structure, which was intentionally chosen not to be among the most stable ($G_8\text{-compact}$, see Table 1). While the trajectories obtained at different temperatures showed the potential for sampling other areas of the potential energy surface and did locate new low-energy structures, the AMOEBA force field cannot describe the breaking or formation of covalent bonds and, therefore, SB structures could not be sampled. The REMD simulations carried out in the present work were started from both CS and SB structures, using appropriate parameters for the different functional groups. The efficiency of the REMD computational strategy is illustrated in Figure 5, where we show the energy of the replica at 274 K in the first REMD iteration of SB structures, after local minimization with the AMOEBA force field as a function of simulation time. For comparison, the minimized energy obtained from a regular MD trajectory at the same temperature and initialized from the same configuration is also displayed. Clearly, the set of isomers sampled by regular molecular dynamics (without exchanges) is rather limited, indicating that this MD trajectory was stuck in metastable regions of the landscape.

The REMD sampling stage provided large databases of candidate structures, from which about 40, all lying in a 5 kcal/mol window, were chosen for further

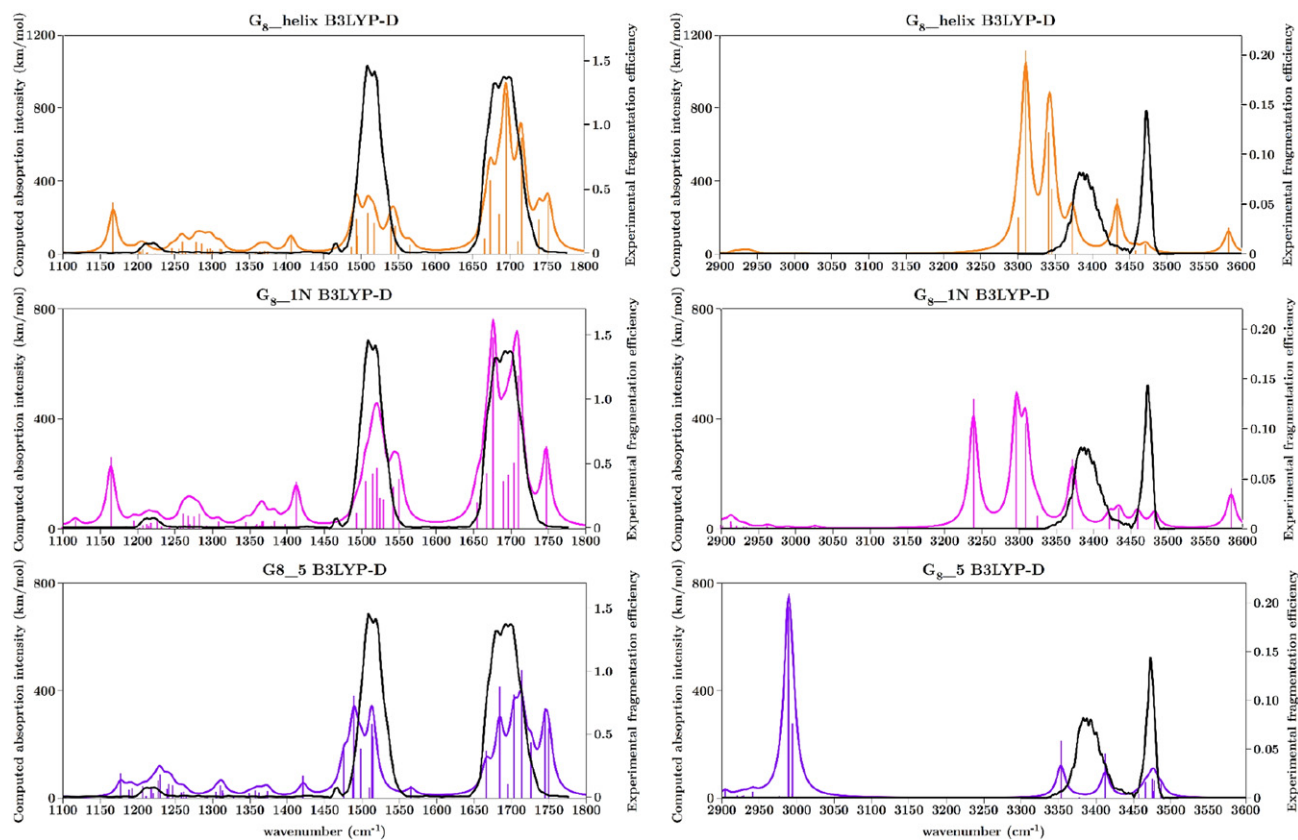


Figure 2. IRMPD spectra of $G_8\text{-Na}^+$ and computed IR absorption spectra for the $G_8\text{-helix}$, $G_8\text{-1N}$, and $G_8\text{-5}$ structures at the B3LYP-D/SVP level using a scaling factor of 0.96.

re-optimization by quantum chemistry methods. Structural refinement generated a significant number of duplicates, including mirror image pairs, which were eliminated. Structures were considered to be duplicates when the root mean square difference of their cartesian structures was smaller than 0.1 Å. In the course of DFT-D re-optimization, some of the SB structures generated with AMOEBA collapsed to CS geometries. Since such rearrangements involve proton transfer, the resulting structures all have the C-terminus O–H in strong hydrogen bonding interaction. The final energetics confirmed the adequacy of the AMOEBA force field to describe $G_8\text{-Na}^+$ as the energy orders were nearly identical. It is noteworthy that most of these low-energy structures escaped initial searches based on chemical intuition.

The description below uses the sequence numbering from the N- to the C-terminus with superscripts, i.e., NH^2 and CO^2 stand for the peptidic N–H and C=O bonds of the second residue. Special cases are for residue 1 where the N–H bonds are in a NH_2 rather than an amide group and for residue 8 for which CO^8 is an acidic rather than an amide C=O. Residue 8 also bears the single O–H bond of $G_8\text{-Na}^+$.

The final energetic results are given in Table 1 at the B3LYP-D//B3LYP-D and MP2//B3LYP-D levels, except for SB_1 and CS_1 for which MP2//MP2 results

are also available. The first five structures in Table 1 have been described previously [16]. Their names have been kept the same for clarity. Other structures (CS_1–CS_10 and SB_1–SB_5) are new and are denoted by their CS or SB type plus their number in MP2//B3LYP-D free-energy order within the CS and SB families. Inspection of Table 1 indicates that B3LYP-D//B3LYP-D and MP2//B3LYP-D relative energies are in very good agreement. Most values are within 1 kcal/mol of one another and the largest discrepancies are ca. 3 kcal/mol. The same holds true for free energies since thermal and entropy contributions are based on B3LYP-D geometries and frequencies in both cases. The stability order is only slightly different when comparing B3LYP-D//B3LYP-D and MP2//B3LYP-D results, as well as when comparing relative energies and free energies at a given computational level, so that differences will not be commented further. The discussion below will be based on MP2//B3LYP-D free energies unless otherwise indicated.

Elimination of Structural Families on the Basis of Their Computed Spectra

The structures that were subjected to full quantum chemical study are widely different according to the

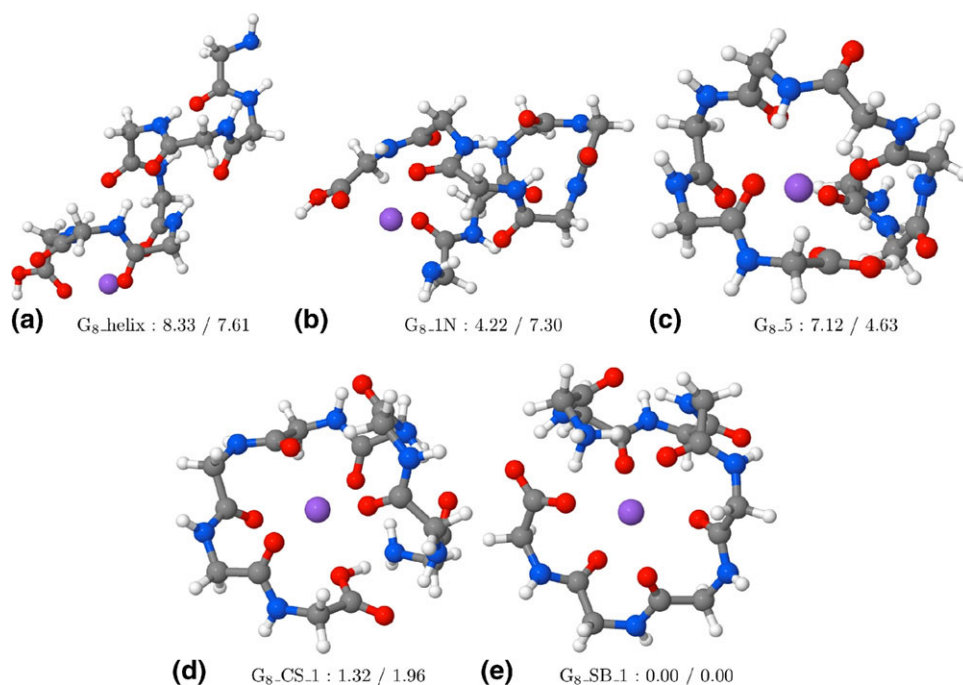


Figure 3. $G_8\text{-Na}^+$ optimized structures (see text for labeling). Relative MP2//B3LYP energies in kcal/mol are reported as electronic energies/Gibbs free energies.

number of carbonyl oxygens attached to sodium (with Na–O distances of 2.7 Å or less), which varies from 3 to 6, and to the number and strength of hydrogen bonds. Some of them may be easily shown not to be good candidates for the experimentally observed $G_8\text{-Na}^+$ ions, based on the comparison of the computed and experimental spectra. In this section we describe some of these cases, illustrating how the final set of structures was selected.

The most stable α -helical structure $G_8\text{-helix}$ is shown in Figure 3. The sodium is bound at the C-terminus and interacts with CO⁵, CO⁶, and CO⁸. It has the C-terminus OH free of any interaction, leading to O–H stretch and COH bend peaks at 3583 and 1167 cm^{-1} , respectively, which are both absent from the IRMPD spectrum (see Figure 2). The free O–H stretch has been previously identified for smaller protonated or cationized amino acids and peptides [33–36]. The COH bend has been used as a diagnostic tool as well [15, 37–42]. Among peptidic N–Hs, NH⁵ interacts weakly with CO¹ with an amide A (N–H stretching) frequency at 3433 cm^{-1} , lying just in between the experimental bands. Furthermore, the N–H—O=C hydrogen bonding pattern typical of a helix leads to a group of amide A frequencies that are stronger than that of NH⁵, in agreement with those of the helical peptides Ac-FA_{*n*}-K-H⁺ ($n = 5, 10$) in the 3300–3500 cm^{-1} region [36], and thus are red-shifted in the present case compared with the IRMPD feature at 3340–3440 cm^{-1} by ca. 70 cm^{-1} . Clearly, the α -helical structure may be excluded for $G_8\text{-Na}^+$. This is in agreement with its computed free-energy of 7.6 kcal/mol relative to that of the most stable structure (see Table 1).

Structure $G_8\text{-1N}$ is another case that can be easily eliminated. Although very different from the α -helix, it also has a free OH, which is incompatible with the IRMPD spectrum. Some of the amide A bands are even more red-shifted than for the α -helix, leading to more inconsistency with IRMPD. Its relative free-energy of $G_8\text{-1N}$ of 7.3 kcal/mol is also unfavorable, leading us to discard it.

A third example is structure $G_8\text{-5}$, in which there are five peptidic N–Hs with no hydrogen bonds, with amide A frequencies between 3466 and 3487 cm^{-1} , in good agreement with the IRMPD band at 3460–3490 cm^{-1} . There are also two bands, the N–H⁵ stretch (involved in a C₁₁ motif with CO⁷) and the symmetric NH₂ stretch (involved in a C₅ motif with NH²) at 3412 and 3378 cm^{-1} , respectively, which are in good agreement with the IRMPD massif between 3350 and 3440 cm^{-1} . This would lead to the narrow experimental feature being the convolution of five bands, while the broader feature would encompass only two, which may not be realistic. The carboxylic O–H is hydrogen-bound to the CO², accounting for the lack of IRMPD bands near 1160 and 3550 cm^{-1} discussed above. However, this hydrogen bond is strong enough that the O–H stretch is computed to be red-shifted to 2989 cm^{-1} , a band that is clearly absent from the IRMPD spectrum. In addition, the amide I massif is computed to be too wide with two bands near 1750 cm^{-1} , which are clearly at odds with the experimental result.

The previous discussion shows that although the resolution of the IRMPD spectrum recorded at room temperature is low, band positions carry enough information to make a severe selection among low-energy

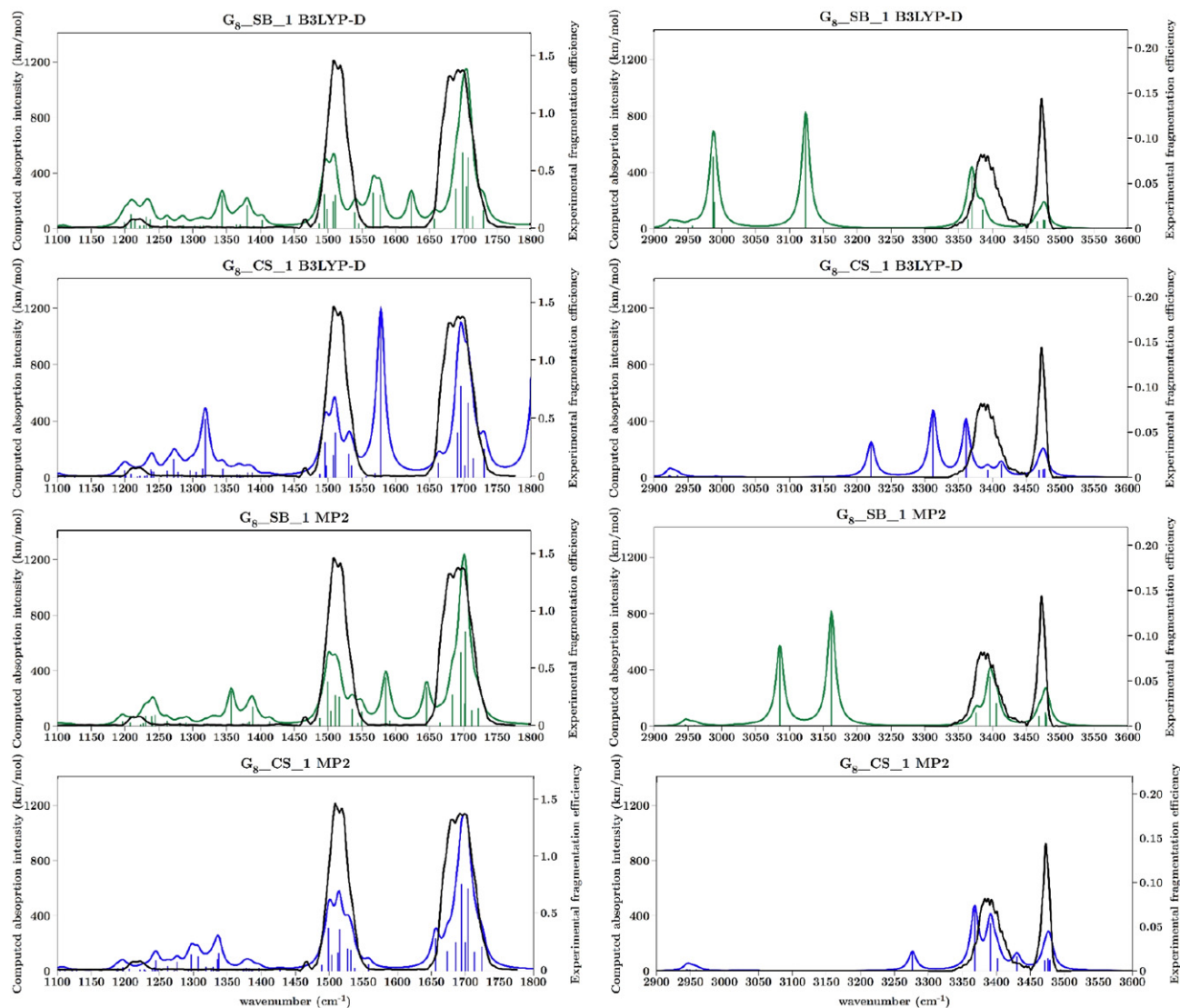


Figure 4. IRMPD spectra of $G_8\text{-Na}^+$ and computed IR absorption spectra for the CS_1 and SB_1 low-energy structures at the B3LYP-D/SVP level using a scaling factor of 0.96, and at MP2/SVP level using a scaling factor of 0.943.

structures. In the following section, we compare the experimental IRMPD spectrum with spectra computed for the two most stable structures. Both satisfy the criteria established above: lack of free O–H and adequate distribution of hydrogen bonds for peptidic N–Hs. The IR spectra of some of the other structures computed at the B3LYP-D level are fairly similar. Owing to the rather high computational level of the present investigation, the lowest-energy structures obtained in this work should be the most abundantly populated, if not exclusively formed in the experiment, although experimental trapping of gaseous ions into local minima cannot be ruled out completely. The two most stable structures happen to be one CS and one SB isomer, which are separated from the others by a significant energy gap (especially for the SB series, see Table 1).

Comparison of IRMPD and DFT Computed Spectra for the Two Lowest Energy Structures

The most stable SB and CS structures are shown in Figure 3, and their computed spectra are shown in Figure 4. We focus on B3LYP-D spectra in this section. The two structures are very similar to each other, with the same set of carbonyl oxygens bound to sodium and the same network of hydrogen bonds except, of course, for the two termini. The peptidic bonds NH^3 , NH^6 , NH^7 , and NH^8 are free, leading in both cases to four bands in the $3468\text{--}3478\text{ cm}^{-1}$ range, in good agreement with the IRMPD feature at $3460\text{--}3490\text{ cm}^{-1}$. The other bands in this region correspond to H-bound peptidic N–Hs and to amine and ammonium bands in CS_1 and SB_1, respectively. Some of them fit reasonably well with the IRMPD broad band at $3340\text{--}3440$

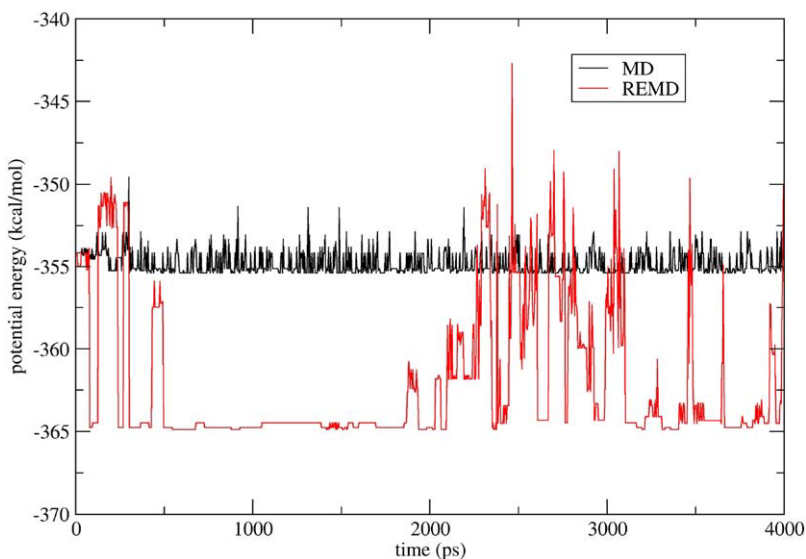


Figure 5. Potential energy obtained after local re-optimization during the first iteration of REMD (in red) and regular MD (in black) simulations of the SB isomer at 274 K, as a function of time. The two simulations were initialized from the same configuration.

cm^{-1} (three in each case), while others appear at significantly lower frequencies. In the SB₁ isomer, the ammonium group has one free N–H bond (with a computed frequency of 3364 cm^{-1}), one bound to CO^4 (N–H–O=C of 1.79 \AA , N–H stretching frequency of 2980 cm^{-1}), and the third bound to an oxygen of the carboxylate at the C terminus with a very small N–H–O=C distance of 1.44 \AA . The latter N–H bond is significantly elongated at 1.13 \AA . This unusually strong hydrogen bond, together with the electrostatic attraction between the ammonium and the carboxylate, leads to a massive red-shift of the N–H stretch frequency to 1890 cm^{-1} . In the CS₁ isomer, one N–H of the N-terminus NH_2 is free while the other is bound to CO^4 although with a distance of 2.0 \AA that is larger than in SB₁ as expected for a non-ionic hydrogen bond. The asymmetric and symmetric NH_2 stretch frequencies are computed at 3393 and 3221 cm^{-1} , respectively. The C-terminus O–H bond interacts strongly with the amine nitrogen (O–H of 1.09 \AA , O–H–N hydrogen bond of 1.47 \AA), leading here again to a massive red-shift of the O–H stretch frequency to 1809 cm^{-1} . The proximity of this frequency with that of the carboxylic C=O and their close interaction lead to a strong out-of-phase mixing in the 1809 cm^{-1} band, and to the appearance of a band corresponding to their in-phase mixing at 1578 cm^{-1} . This band emerges between the amide I and II massifs, a feature that is absent from the IRMPD spectrum. In the same region, the SB₁ isomer displays three HNH bending modes of the ammonium group at 1569 , 1577 , and 1624 cm^{-1} . Clearly, both cases have many features in good agreement with the IRMPD spectrum, yet significant discrepancies between theory and experiment also exist in both spectral ranges, such that attribution of the spectrum to either structure is not possible.

Comparison of IRMPD and MP2 Computed Spectra for the Two Lowest Energy Structures

The previous discrepancies may also occur because the nonbonded interactions, which are present in the folded structures described above, are not well described by the B3LYP functional complemented by the empirical correction used because the ionic hydrogen bonds are particularly strong in the present case. To improve the computational level, we repeated the geometry optimizations and frequency calculations of SB₁ and CS₁ at the RI-MP2/SVP level. Both minima could be located at this level as well, with some structural differences. In SB₁, the MP2 N–H/NH–O distances are $1.12/1.45\text{ \AA}$ versus $1.13/1.44\text{ \AA}$ with B3LYP-D. In CS₁, the MP2 O–H/OH–N distances are $1.07/1.51\text{ \AA}$, while they are $1.09/1.47\text{ \AA}$ with B3LYP-D. Although small, these differences are not negligible. In all cases, the MP2 covalent bond lengths are less elongated, and the hydrogen bonding distances are larger than at the B3LYP-D level. Since N–H and O–H stretching frequencies are strongly sensitive to hydrogen bonding, it is not surprising that such changes lead to important differences in the computed spectra, inducing a series of blue shifts in band positions for the corresponding stretching modes. The MP2//MP2 free-energy difference between the two structures is negligible at 0.49 kcal/mol . The computed IR spectra of CS₁ and SB₁ at the B3LYP-D and MP2 levels are shown in Figure 4. The IR spectrum of SB₁ in the fingerprint region is not much different at the MP2 and B3LYP-D levels, except that the ammonium N–H band is shifted from 1890 to 1947 cm^{-1} . Because of limitations on the electron energy range, IRMPD experiments are not usually carried out in the high-energy end of the fingerprint region. We did scan the $1900\text{--}2400\text{ cm}^{-1}$ range in a pass where the laser power was

nearly constant at 0.6 W. No band was detected in this region. However, such extremely strong hydrogen bonding is expected to be associated to large anharmonicity such that harmonic computations are not expected to be particularly reliable, as discussed recently for deprotonated glutamic and aspartic acids [37].

While the amide I and amide II broad bands are well positioned in both cases, blue shifts of the HNH bending modes of the ammonium at 1585, 1591, and 1646 cm^{-1} still lead to bands in between the amide bands that are not observed experimentally. In the high-frequency region, the three amide A frequencies of H-bound N–Hs are slightly blue-shifted such that they agree very well with the IRMPD band at 3340–3440 cm^{-1} . However the ammonium N–H stretches, although significantly blue-shifted, still show exceedingly low frequencies with respect to experiment. Analogously large red shifts have been previously described for N–H stretches of ammonium groups that are hydrogen-bonded to carbonyl oxygens, with bands in the 3000–3200 cm^{-1} area [33, 35]. Careful re-scanning after laser power optimization in this region confirmed the absence of any experimental band.

Blue shifts are also observed in the MP2 versus the B3LYP-D spectrum of CS₁. The O–H stretch frequency is displaced from 1809 to 1950 cm^{-1} , without significant coupling with the carboxylic C=O stretch and disappearance of the band at 1578 cm^{-1} . Again, there is no equivalent band in the experimental spectrum. The carboxylic C=O stretch at 1656 cm^{-1} is within the amide I massif, while the NH₂ bend at 1558 cm^{-1} is within the amide II. Thus, experiments and computations match very well in the fingerprint region. Amide A bands of the two strongly H-bound N–Hs are blue-shifted by ca. 50 cm^{-1} , such that they fit well with the experimental band at 3340–3440 cm^{-1} . The symmetric NH₂ stretch, although shifted by 50 cm^{-1} as well, is the only band not seen experimentally. Its weak intensity is probably responsible for this lack of observation. Overall, the MP2 spectrum of CS₁ is in very good agreement with the IRMPD spectrum. One may then conclude that this structure is the most likely candidate for the room-temperature structure of G₈-Na⁺.

Discussion

To characterize these very strong interactions in more detail, the transition-state for interconversion of CS₁ and SB₁, i.e., proton transfer between the N- and C-termini was determined at the B3LYP-D/SVP level. It was found to be very similar to the SB₁ and CS₁ minima it connects, with N–H and O–H distances of 1.310 and 1.201 Å, respectively. The activation barrier is computed to be 0.3 and –0.06 kcal/mol at the B3LYP-D/B3LYP-D and MP2//B3LYP-D levels, respectively, relative to the CS₁ minimum. Thus, one would expect that easy proton transfer takes place and that the overall structure may not be described properly as either SB₁ or CS₁. Our results indicate, however, that the com-

puted spectrum of CS₁ fits well with the experimental spectrum.

The salt bridge versus charge solvation structures of cationized amino acids have been a matter of intense research using IRMPD spectroscopy in the last few years [38]. This has been extended to cationized peptides recently [38, 39, 41–43]. Most cases of cationized peptides for which a SB structure was established contain arginine, a highly basic residue that favors proton transfer [38, 39, 41]. Only when a divalent ion was attached to the AA dipeptide was a SB structure found to be favored, while the alkali cationized di- and tripeptides favored a CS isomer [15, 42, 43]. Apparently, the present results are the first to point to a high stability for a salt bridge in a cationized peptide without any basic residue and limited electrostatic perturbation induced by an alkali cation.

Our results indicate that there may be no general way to distinguish between CS and SB isomers in peptides with more than a few residues on the basis of their IRMPD spectra, especially in the fingerprint region. While carboxylic O–H bonds that are free of hydrogen bonding are simple to detect (O–H stretching and COH bending frequencies near 3550 and 1160 cm^{-1} , respectively, plus carboxylic C=O stretching frequency near or above 1750 cm^{-1}), it is unlikely that a carboxylic acid remains free of interaction in peptides of growing size. Any carboxylic acid involved in strong hydrogen bonding with, e.g., an amine, may be rather difficult to distinguish from its isomeric carboxylate in strong hydrogen bonding with, e.g., an ammonium, especially when the growing size of the peptide makes the amide I and II massifs grow broader. This is illustrated by the very similar amide I and II massifs computed for the CS₁ and SB₁ isomers of G₈-Na⁺ (see Figure 4). Some of the ammonium bending bands appear near 1400 cm^{-1} when free but are blue-shifted by hydrogen bonding [42] and may then be hidden, at least partially, by the amide II massif. Our results point to the possibility that ammonium N–H stretches appear below ca. 3300 cm^{-1} as a result of strong, ionic hydrogen bonding. Such bands would then be typical of salt bridges, assuming that CS isomers cannot give rise to analogous shifts. Firm conclusions along these lines await more comprehensive experimental results on ionic peptides. In any event, the present results indicate that although G₈-Na⁺ appears to have a CS structure, salt bridges may be highly stable even for peptides of growing size in which the metal charge can be efficiently solvated by carbonyl oxygens, even without basic residues. Dynamical studies would be particularly welcome on such systems with a shared proton [44], however, current limitations of density functionals (see below) make such calculations of limited use at least in the present case.

Although B3LYP-D provides energetics in very good agreement with MP2, it appears to yield structures that are slightly too compact, generating significant errors on IR band positions whenever there are strong hydro-

gen bonds. In such cases, red shifts of ca. 50 cm⁻¹ with respect to MP2 values have been obtained for G₈-Na⁺. This trend indicates that some of the results reported may not be free of similar errors. More work is needed to establish a reliable quantum chemical level of calculation that is able to yield reliable IR spectra for still larger biomolecules. Since vibrational frequency calculations cannot be carried out at the MP2 level for peptides much larger than G₈, the most promising track remains DFT now. However, the adequate density functional for treating charged molecules with very strong hydrogen bonds remains to be identified. Work is also underway on oligoglycines of different sizes to establish if the structures probed experimentally are of the charge solvation or salt bridge type and to understand their dynamical behavior in more detail.

Conclusions

IRMPD spectra have been obtained for G₈-Na⁺ at room temperature in the 1000–1900 and 3000–3700 cm⁻¹ regions. The results indicate that the C-terminus OH group is either hydrogen bound or absent. The 3000–3700 cm⁻¹ spectrum is indicative of a network of N–H bonds, most of which are either free or not very strongly hydrogen-bound. To gain more structural insight, extensive molecular modeling was carried out. A powerful approach coupling the polarizable force field AMOEBA and replica exchange molecular dynamics was used to explore the conformation landscape of G₈-Na⁺ efficiently. Both charge solvation and salt bridge structures were identified and refined at the B3LYP-D and MP2 quantum chemical levels. It turned out that the two structures with the most favorable free energies at 298 K are a CS/SB pair related by a simple proton transfer between the N- and C-termini with a high degree of peptidic C=O coordination to the sodium. The activation barrier for interconversion of these structures was computed to be quite small. It was only at the MP2 level that good agreement between experiments and computations could be obtained, yielding a charge solvation structure as the most likely candidate for the gas-phase structure of G₈-Na⁺ at 298 K.

Acknowledgments

O.P.B. thanks Ecole Polytechnique for a 1 year post-doctoral fellowship and the European Union (EU) commission for a 1 year post-doctoral support through the STREP 15,637 “EPITOPES” funded by the New and Emerging Science and Technology program of the 6th pluriannual framework of EU. C.F.C. thanks the Fundação para a Ciência e a Tecnologia for a 2-year post-doctoral fellowship. Implementation of the REMD method was supported by a travel grant to D.S. and C.C. from GDR 2758 funded by CNRS. This work was granted access to the HPC resources of [CCRT/CINES/IDRIS] under the allocation x2009085107 made by GENCI (Grand Equipement National de Calcul Intensif).

References

1. Michell A. R., Ed. *The Clinical Biology of Sodium: The Physiology and Pathophysiology of Sodium in Mammals*; Pergamon: Oxford, UK, 1995; p. 16.
2. Skou, J. C.; Norby, J. G.; Maunsbach, A. B.; Esmann, M. In *Progress in Clinical and Biological Research*; Vol. CCLXVIII; Alan R. Liss, Inc.: New York, 1987; pp 401–632.
3. Aidley, D. J.; Stanfield, P. R. *Ion Channels: Molecules in Action*; Cambridge University Press: Cambridge, 1994, pp 1–58.
4. Teesch, L. M.; Orlando, R. C.; Adams, J. Location of the Alkali Metal Ion in Gas-Phase Peptide Complexes. *J. Am. Chem. Soc.* **1991**, *113*, 3668–2675.
5. Lin, T.; Payne, A. H.; Glish, G. L. Dissociation Pathways of Alkali-Cationized Peptides: Opportunities for C-Terminal Peptide Sequencing. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 497–504.
6. Feng, W. Y.; Gronert, S.; Fletcher, K. A.; Warres, A.; Lebrilla, C. B. The Mechanism of C-Terminal Fragmentations in Alkali Metal Ion Complexes of Peptides. *Int. J. Mass Spectrom.* **2003**, *222*, 117–134.
7. Bulleigh, K.; Howard, A.; Do, T.; Wu, Q.; Anbalagan, V.; van Stipdonk, M. Investigation of Intramolecular Proton Migration in a Series of Model, Metal-Cationized Tripeptides Using In Situ Generation of an Isotope Label. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 227–232.
8. Lee, S. W.; Kim, H. S.; Beauchamp, J. L. Salt Bridge Chemistry Applied to Gas-Phase Peptide Sequencing: Selective Fragmentation of Sodiated Peptide Ions Adjacent to Aspartic Acid Residues. *J. Am. Chem. Soc.* **1998**, *120*, 3188–3195.
9. Wyttenbach, T.; Bushnell, J. E.; Bowers, M. T. Salt Bridge Structures in the Absence of Solvent? The Case for the Oligoglycines. *J. Am. Chem. Soc.* **1998**, *120*, 5098–5103.
10. Kohtani, M.; Jarrold, M. F.; Wee, S.; O’Hair, R. A. J. Metal-Ion Interactions with Polyalanine Peptides. *J. Phys. Chem. B* **2004**, *108*, 6093–6097.
11. Cox, H. A.; Julian, R. R.; Lee, S.-W.; Beauchamp, J. L. Gas-Phase H/D Exchange of Sodiated Glycine Oligomers with ND₃: Exchange Kinetics Do Not Reflect Parent Ion Structures. *J. Am. Chem. Soc.* **2004**, *126*, 6485–6490.
12. Kish, M. M.; Wesdemiotis, C.; Ohanessian, G. The Sodium Ion Affinity of Glycylglycine. *J. Phys. Chem. B* **2004**, *108*, 3086–3091.
13. Wang, P.; Kapota, C.; Ohanessian, G.; Wesdemiotis, C. The Sodium Ion Affinities of Simple Di-, Tri-, and Tetrapeptides. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 541–552.
14. Ye, S. J.; Armentrout, P. B. Absolute Thermodynamic Measurements of Alkali Metal Cation Interactions with a Simple Dipeptide and Tripeptide. *J. Phys. Chem. A* **2008**, *112*, 3587–3596.
15. Balaj, O. P.; Kapota, C.; Lemaire, J.; Ohanessian, G. Vibrational Signatures of Sodiated Oligopeptides (GG-Na⁺, AA-Na⁺, GGG-Na⁺, and AAA-Na⁺) in the Gas Phase. *Int. J. Mass Spectrom.* **2008**, *269*, 196–209.
16. Semrouni, D.; Ohanessian, G.; Clavaguera, C. Structural, Energetic, and Dynamical Properties of Sodiated Oligoglycines: Relevance of a Polarizable Force Field. *Phys. Chem. Chem. Phys.* **2010**, DOI: 10.1039/b924317h.
17. Ren, P. Y.; Ponder, J. W. Consistent Treatment of Inter- and Intramolecular Polarization in Molecular Mechanic Calculations. *J. Comput. Chem.* **2002**, *23*, 1497–1506.
18. Grossfield, A.; Ren, P. Y.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *J. Am. Chem. Soc.* **2003**, *125*, 15671–15682.
19. Voelz, V. A.; Shell, M. S.; Dill, K. A. Predicting Peptide Structures in Native Proteins from Physical Simulations of Fragments. *PLoS Comp. Biol.* **2009**, *5*(2), e1000281.
20. Buchete, N. V.; Hummer, G. Protein Folding Kinetics from Replica Exchange Molecular Dynamics. *Phys. Rev. E* **2008**, *77*(3), 030902.
21. Weinstock, D. S.; Narayanan, C.; Felts, A. K.; Andrec, M.; Levy, R. M.; Wu, K. P.; Baum, J. Distinguishing Between Structural Ensembles of the GB1 Peptides: REMD Simulations and NMR Experiments. *J. Am. Chem. Soc.* **2007**, *129*, 4858–4859.
22. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
23. Ponder, J. W. TINKER, Software Tools for Molecular Design, Version 4.2, **2004**.
24. Penev, E. S.; Lampoudi, S.; Shea, J.-E. TiReX: Replica-Exchange Molecular Dynamics Using TINKER. *Comp. Phys. Commun.* **2009**, *180*, 2013–2019.
25. Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
26. Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
27. Bouteiller, Y.; Pouilly, J.-C.; Desfrancois, C.; Grégoire, G. Evaluation of MP2, DFT, and DFT-D Methods for the Prediction of Infrared Spectra of Peptides. *J. Phys. Chem. A* **2009**, *113*, 6301–6307.
28. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic Structure Calculations on Workstation Computers: The Program System Turbomole. *Chem. Phys. Lett.* **1989**, *162*, 165–169. For the current version, see <http://www.turbomole.com>.
29. MacAleese, L.; Simon, A.; McMahon, T. B.; Ortega, J. M.; Scuderi, D.; Lemaire, J.; Maitre, P. Mid-IR Spectroscopy of Protonated Leucine

- Methyl Ester Performed with an FTICR or a Paul Type Ion-Trap. *Int. J. Mass Spectrom.* **2006**, *249*, 14–20.
30. MacAleese, L.; Maitre, P. Infrared Spectroscopy of Organometallic Ions in the Gas Phase: From Model to Real World Complexes. *Mass Spectrom. Rev.* **2007**, *26*, 583–605.
 31. Bakker, J. M.; Besson, T.; Lemaire, J.; Scuderi, D.; Maitre, P. Gas-Phase Structure of a π -Allyl-Palladium Complex: Efficient Infrared Spectroscopy in a 7 T Fourier Transform Mass Spectrometer. *J. Phys. Chem. A* **2007**, *111*, 13415–13424.
 32. Bakker, J. M.; Sinha, R. K.; Besson, T.; Brugnara, M.; Tosi, P.; Salpin, J. Y.; Maitre, P. Tautomerism of Uracil Probed Via Infrared Spectroscopy of Singly Hydrated Protonated Uracil. *J. Phys. Chem. A* **2008**, *112*, 12393–12400.
 33. Stearns, J. A.; Mercier, S.; Seaiby, C.; Guidi, M.; Boyarkin, O.; Rizzo, T. R. Conformation-Specific Spectroscopy and Photodissociation of Cold, Protonated Tyrosine and Phenylalanine. *J. Am. Chem. Soc.* **2007**, *129*, 11814–11820.
 34. Bush, M. F.; O'Brien, J. T.; Prell, J. S.; Saykally, R. J.; Williams, E. R. Infrared Spectroscopy of Cationized Arginine in the Gas Phase: Direct Evidence for the Transition from Nonzwitterionic to Zwitterionic Structure. *J. Am. Chem. Soc.* **2007**, *129*, 1612–1622.
 35. Vaden, T. D.; de Boer, T. S. J. A.; Simons, J. P.; Snoek, L. C.; Suhai, S.; Paizs, B. Vibrational Spectroscopy and Conformational Structure of Protonated Polyalanine Peptides Isolated in the Gas Phase. *J. Phys. Chem. A* **2008**, *112*, 4608–4616.
 36. Stearns, J. A.; Seaiby, C.; Boyarkin, O.; Rizzo, T. R. Spectroscopy and Conformational Preferences of Gas-Phase Helices. *Phys. Chem., Chem. Phys.* **2009**, *11*, 125–132.
 37. Oomens, J.; Steill, J. D.; Redlich, B. Gas-Phase IR Spectroscopy of Deprotonated Amino Acids. *J. Am. Chem. Soc.* **2009**, *131*, 4310–4319.
 38. Prell, J. S.; O'Brien, J. T.; Steill, J. D.; Oomens, J.; Williams, E. R. Structures of Protonated Dipeptides: The Role of Arginine in Stabilizing Salt Bridges. *J. Am. Chem. Soc.* **2009**, *131*, 11442–11449.
 39. Prell, J. S.; Demireva, M.; Oomens, J.; Williams, E. R. Role of Sequence in Salt-Bridge Formation for Alkali Metal Cationized GlyArg and ArgGly Investigated with IRMPD Spectroscopy and Theory. *J. Am. Chem. Soc.* **2009**, *131*, 1232–1242.
 40. Polfer, N. C.; Oomens, J. Vibrational Spectroscopy of Bare and Solvated Ionic Complexes of Biological Relevance. *Mass Spectrom. Rev.* **2009**, *28*, 468–494.
 41. Polfer, N. C.; Paizs, B.; Snoek, L. C.; Compagnon, I.; Suhai, S.; Meijer, G.; von Helden, G.; Oomens, J. Infrared Fingerprint Spectroscopy and Theoretical Studies of Potassium Ion Tagged Amino Acids and Peptides in the Gas Phase. *J. Am. Chem. Soc.* **2005**, *127*, 8571–8579.
 42. Polfer, N. C.; Oomens, J.; Dunbar, R. C. Alkali Metal Complexes of the Dipeptides PheAla and AlaPhe: IRMPD Spectroscopy. *Chem. Phys. Chem.* **2008**, *9*, 579–589.
 43. Dunbar, R. C.; Steill, J. D.; Polfer, N. C.; Oomens, J. Peptide Length, Steric Effects, and Ion Solvation Govern Zwitterion Stabilization in Barium-Chelated Di- and Tripeptides Dunbar. *J. Phys. Chem. B* **2009**, *113*, 10552–10554.
 44. Cimas, A.; Vaden, T. D.; de Boer, T. S. J. A.; Snoek, L. C.; Gaigeot, M. P. Vibrational Spectra of Small Protonated Peptides from Finite Temperature MD Simulations and IRMPD Spectroscopy. *J. Chem. Theor. Comput.* **2009**, *5*, 1068–1078.